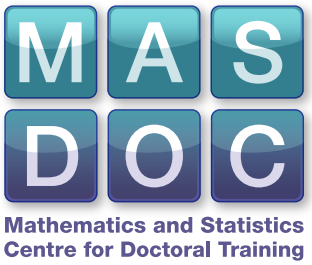


Quasi-Stationary Distributions for Evolving Epidemic Models: Simulation and Characterisation

by

Adam Griffin

Thesis

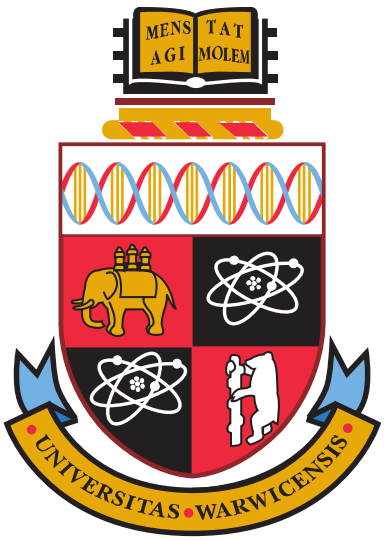
Submitted for the degree of

Doctor of Philosophy

Mathematics Institute, Statistics Department

The University of Warwick

September 2016



Contents

List of Figures

Acknowledgments

I would like to thank my supervisors Simon and Gareth for their constant encouragement and words of wisdom. I would also like to thank all my friends and my partner for their tireless patience.

The author's research is funded through MASDOC DTC by EPSRC: Grant Number EP/HO23364/1.

Declarations

This text is wholly the work of the author except for Chapter 2, which is a literature review, and any cited results in Chapter 3. Chapter 5 has been published as a paper submitted to Applied Probability Press, "Simulation from Quasi-Stationary Distributions on Reducible State Spaces" with co-authors P.A. Jenkins, G.O. Roberts and S.E.F. Spencer, but the author of this thesis was the primary author of the paper.

This thesis, or any part of it, has not been submitted as part of a degree at this, or any, university.

Abstract

This thesis develops probabilistic models for the spread of infectious diseases in which individuals experience a period of transient immunity after recovering from infection. Quasi-stationary distributions (QSDs) and limiting conditional distributions (LCDs) are used to describe the temporary equilibrium that is reached between an initial exponential growth phase and the epidemic dying out. This thesis includes results characterising QSDs corresponding to existing birth-death processes and epidemic models and to new processes such as the Evolving Strain SIRS model which we define to describe the progression of a disease undergoing antigenic drift, such as seasonal influenza. Existence and uniqueness results are proven for specific LCDs. Results regarding marginals of special cases of these processes are proven, including the preservation of x-invariance for Q as discussed in Pollett [1988].

Many of the models considered in this thesis are multidimensional, which makes explicit calculation of QSDs extremely challenging. To combat this, specialised techniques for simulating QSDs are developed to illustrate and explore these distributions. These novel methods, involving variants on SMC samplers, are shown to facilitate the simulation of QSDs for discrete-valued stochastic processes, particularly reducible processes. A formal proof of convergence of the SMC sampler is provided for some simple examples.

The simulation methods are then used to characterise the properties of QSDs and LCDs related to endemic epidemic models with evolving strains under an equivalence relation. These QSDs are used to define a reproduction number similar to R_0 when the process starts from quasi-stationarity. The epidemic models with evolving strains are shown to have the standard SIR and SIRS epidemic models arising as limiting processes as evolution at each infection becomes certain.

Abbreviations

α	Decay Parameter
α_0	Absorption Parameter
BDP	Birth-Death Process
E-BDP	Birth-Death Process with Evolving Strains
E-SIS	SIS Model with Evolving Strains
E-SIRS	SIRS Model with Evolving Strains
E-TI	Transient Immunity Process with Evolving Strains
LCD	Limiting Conditional Distribution
p.g.f.	probability generating function
QSD	Quasi-Stationary Distribution
SMC	Sequential Monte Carlo
R_0	Basic Reproductive Number
R_H	Household Reproductive Number
R_Q	Quasi-Stationary Reproductive Number
R_*	Modified Household Reproductive Number

Chapter 1

Introduction

The aim of this thesis is to develop probabilistic models which represent the levels of immunity in populations experiencing the circulation of endemic diseases such as influenza, and from that draw conclusions about features of the model. To this end we will make use of quasi-stationary distributions to describe the behaviour of such models before the end of the epidemic.

We consider three main facets of a central theme. Firstly, existence and uniqueness of quasi-stationary distributions (QSDs) for endemic diseases is considered, along with observing characteristics of these distributions. Secondly, new endemic disease models are developed and analysed; these are related to existing work in the field. These models incorporate the notions of transient and permanent immunity within a population, and link it to the evolution of pathogens within a population. Thirdly, new simulation techniques involving Sequential Monte Carlo resampling methods are developed in order to draw from QSDs for these models, particularly when they are reducible processes.

Quasi-Stationary Distributions for Endemic Diseases

When considering epidemic models to describe endemic diseases, such as influenza, we note that many epidemic models go through an initial peak, and then die off. However, we see that real diseases stay in the population for a very long time, before some intervention or some rare event causes them to die out. Before this, many such

diseases experience a steady-state behaviour, which leads us to consider conditioning epidemic models on non-extinction. This conditional equilibrium is formalised as a quasi-stationarity distribution.

Quasi-stationary distributions, defined in Chapter 2.1, date back to work by Yaglom [1947] on limit theorems and work by Wright [1931] into gene frequencies. These QSDs extend the idea of the rare event in which a given process has survived for a long time without hitting some end state, and consider the limiting case where, conditional on not having arrived at this end state, the process has reached some form of pseudo-stability. These Limiting Conditional Distributions (LCDs) were discussed by Mandl [1960]. This can be compared to deterministic systems which reach an endemic equilibrium which often manifests as a fixed point away from the absorbing end-state.

This was continued in work by Cavender [1978] focusing on birth-death processes and in work by Seneta and Vere-Jones [1966] on finite state models. Later this was applied to more complex models examined by Lambert [2008] with regards to genealogical processes, and by Athreya and Ney [1972] on branching processes. Pollett [1999] and van Doorn [1991] have also looked into this in some generality in the case of continuous-time, discrete-space processes. A comprehensive review of the results in this area is given in van Doorn and Pollett [2013]. There also exists a wealth of work on QSDs for diffusion processes, dating back to Mandl [1961], but more recently worked on by Cattiaux et al. [2009]. This is reviewed and extended for general diffusions in Collet et al. [2001], and for population processes in Méléard and Villemonais [2012].

However, many such results lead to analytically intractable quantities or expressions. To this end, many people have looked into approximation and computation techniques which allow one to study the properties of these distributions. With regards to more specific applications to the SIS model Bailey [1975], also known as the logistic epidemic, Kryscio and Lefèvre [1989] have looked into the problem, providing approximating recurrent processes which circumvent the absorbing nature of the SIS model. The efficacy of these approximations was determined in Clancy and Pollett [2003]. Additionally, Nåsell [1999] produced asymptotic results particularly with regards to analytic approximations related to cumulants and moments of the QSDs associated to the SIS model which become exact as the population size goes to infinity. Other approaches include moment closure techniques by Martins et al. [2012] which make use of equations linking the moments of the system, and making

simplifications which make a finite system of solvable equations. Moreover, Blanchet et al. [2013] and Groisman and Jonckheere [2013] have considered simulation techniques related to Monte Carlo methods such as particle filters and Sequential Monte Carlo to simulate LCDs and QSDs.

Epidemic Modelling

The mathematical modelling of epidemics and endemic diseases goes back to Kermack and McKendrick [1927], and the Reed-Frost epidemic model (Abbey [1952]) which presented a compartmental stochastic process to describe the number of infective individuals, which evolved into the standard S-I-R epidemic models which we will define in Section 2.2. This was continued and extended in the seminal work by Bailey [1975]. Andersson and Britton [2000] contains a review of results related to the SIR and Reed-Frost models. In this thesis, the focus is on the levels of immunity within a population, be that permanent immunity to a pathogen or a transient period of immunity before becoming re-susceptible to such a disease. The SIR and SIRS models which describe these two behaviours are considered in particular. Clancy and Mendy [2010] look into population demographics to explain this transient immunity and its effect on endemic diseases. Such epidemic models with demography are also reviewed in Nåsell [2002].

More recently, work has been done to look into modelling more specific diseases which exhibit mutation of the pathogen through methods such as antigenic drift, including influenza and malaria. As a starting point one can consider the Wright-Fisher and Moran models as predecessors to this, which focus on the mutation itself. Saunders [1981] was one of the first to consider multiple co-circulating strains in a population, Girvan et al. [2002] later considered models with a mutation mechanism to move between strains in an individual. Work by Gog and Grenfell [2002] looks at a linear strain space to model diseases similar to influenza using stepwise mutations and a pairwise immunity structure. Parisi et al. [2013] and Bedford et al. [2015] have used more complex immunity and mutation methods to more closely model Influenza A and infer behavioural characteristics of the evolution of the pathogen.

Real-World Example: Influenza

Seasonal influenza is a viral disease which affects 5-10% of the world's population each year and kills around 36 000 people each year in the USA alone. After an individual has been infected by a strain of influenza, the host's immune system begins to create antibodies which attach to HA (haemagglutinin) and NA (neuraminidase) proteins found on the surface of the influenza virus, inhibiting the virus and its ability to attach to cells in the host. Following an infection, the body creates specific antigen receptors to detect the virus. In particular, there are at least 5 key sites on the virus which the antigenic receptors respond to (Parisi et al. [2013]).

However, the influenza virus is constantly mutating, and so the HA proteins may change over time. If the virus becomes sufficiently different from that to which the body's receptors react, then the individual may become infected with this new "strain". Indeed, in a region in which most of a given population have developed an immunity to a given strain, the progeny of the virus will be dominated by those instances where the virus has sufficiently mutated, which in turn lowers the effective immunity of the population. Additionally, due to having had less exposure to various strains, children are more susceptible to contracting and then spreading the newer strains in circulation.

To combat this, the standard preventative measure is the vaccination of at-risk individuals, such as the sick and elderly. Typically, this comes in the form of a trivalent or quadrivalent vaccine which includes versions of three or four strains recommended by the WHO (World Health Organisation [2015]) on a six-monthly basis. However, due to constraints on time to research and manufacture sufficient quantities of vaccines, such decisions need to be made well in advance of a given flu season. This is not straightforward; in 2014-15 the main trivalent vaccine was said by Schotsaert and García-Sastre [2016] to be only 44% effective due to the mutations of the main circulating strains between analysis and deployment. It is hoped that this work and subsequent work aids these decisions.

This thesis aims to take these areas of interest, extending the simple models to incorporate the mutation we see in seasonal influenza, whilst keeping key features which allow us to still find features of the QSDs associated to these new models. In cases where analytic expressions don't exist, new machinery is developed to allow Monte Carlo simulation of the quasi-stationary distributions; these methods have more widespread applications, but QSDs and LCDs are the primary motivation.

Outline of Thesis

Chapter 2 discusses preliminary results related to epidemic models, quasi-stationary distributions and Sequential Monte Carlo which will be made use of in the rest of the thesis, putting the thesis into context within a greater network of research.

Chapter 3 takes existing work on birth-death processes by van Doorn and Pollett [2013] and Collet et al. [2001], and works to characterise the QSDs with infinite mean. We investigate the related generating functions, as well as forming iterative expressions for the higher-energy QSDs. We also investigate the pure death process, characterising the QSDs and giving iterative expressions in some cases.

Chapter 4 introduces the Transient Immunity process. The chapter discusses the behaviour of the process, developing expressions for the generating function and related quantities. From this, it develops two notions of quasi-stationarity: conditioning on infectives still being present in the population, and conditioning on immunity still existing within the population. Behaviour regimes dependent on the relationship of the main parameters are characterised. Finally the behaviour of the process conditional on non-extinction of the immunity is examined.

Chapter 5 focuses on the problems with simulation of quasi-stationary distributions, particularly in the case of infinite or reducible state spaces. The focus is the development of SMC samplers (Del Moral et al. [2006]) to simulate quasi-stationary distributions, with focus on birth-death processes and the Transient Immunity process. New resampling methods are developed – Combine-Split Resampling, and Regional-Resampling – to better explore such QSDs. This is supported by some technical results on convergence in Wasserstein-distance in some simple cases.

Chapter 6 takes the standard epidemic models, along with the Linear Birth-Death Process and the Transient Immunity Process and introduces the notion of the evolution of a pathogen through the use of countably many strains of the pathogen. We will construct these models, and then observe properties about the model through direct calculation and through a simulation study of the models in question. In particular, we look at QSDs of the models and how they are linked between models, and also how the immunity component changes the behaviour of the model.

Concluding remarks and comments on the future of this area of research are made in Chapter 7.

Chapter 2

Preliminary Results

This chapter will outline a number of preliminary concepts and results from other sources which will be made use of in this thesis. We will first outline the concept of quasi-stationarity, and results regarding existence and uniqueness of quasi-stationary distributions in some specific cases with reference to the excellent review paper of van Doorn and Pollett [2013]. Following that, some basic ideas regarding stochastic epidemic models (SIS, SIR, SIRS) will be mentioned regarding construction of such models, and results related to extinction times under different regimes, with reference to Andersson and Britton [2000]. Finally, this chapter will outline the ideas of Sequential Monte Carlo and Sequential Importance Sampling, as developed by Liu and Chen [1998], ahead of their application in simulating quasi-stationary distributions in Chapter 5.

2.1 Definitions for Markov Processes

Let $X = (X(t))_{t\geq 0}$ be a Markov process on a countable state space $\bar{S}$ with infinitesimal rate matrix $\bar{Q} = [q_{ij}]_{i,j\in\bar{S}}$.

Definition 2.1.1. In the rest of this thesis, we will use the notation $\mathbb{P}_i[X(t) \in A]$ to denote the probability of an event under the initial condition $X(0) = i$. Similarly we will use $\mathbb{E}_i[h(X)]$ to be the expectation of a random variable $h(X)$ under the

initial condition $X(0) = i$. More generally we will use the notation

$$\mathbb{P}_\mu[X(t) \in A] := \int_{\bar{S}} \mathbb{P}_x[X(t) \in A]\mu(dx) = \sum_{x \in \bar{S}} \mathbb{P}_x[X(t) \in A]\mu(x) \tag{2.1.1}$$

$$\mathbb{E}_\mu[h(X)] := \int_{\bar{S}} \mathbb{E}_x[h(X)]\mu(dx) = \sum_{x \in \bar{S}} \mathbb{E}_x[h(X)]\mu(x) \tag{2.1.2}$$

for proper probability measures μ over the probability space on which the process of interest is defined. With this we will denote the transition probabilities by $P_{ij}(t) = \mathbb{P}_i[X(t) = j]$ for $i, j \in \bar{S}$, $t \geq 0$.

For all the processes of interest in this thesis defined on a space $\bar{S}$, we denote a partition of $\bar{S}$ into the disjoint union $\bar{S} = S \cup S_0$ where S_0 denotes a "dead" region from which the process cannot leave after entering. We also let $Q = [q_{ij}]_{i,j \in S}$ be the infinitesimal rate matrix restricted to the transient states S. Once X reaches S_0, the process is said to be *absorbed*. In the case of epidemic models and population processes, we may refer to a process having gone *extinct*, but such an absorption event will be clearly defined. We identify the states in S_0 into a single state which we shall denote by 0, and refer to it as an *absorbing state*. Moreover, we shall denote the *absorption time* of the process with initial distribution $\mathbf{v}$ by $T_{\mathbf{v}} = \inf\{t \geq 0 : X(t) = 0 | X(0) \sim \mathbf{v}\}$. When the initial condition is $X(0) = i$ we will use $T_i := \inf\{t \geq 0 : X(t) = 0 | X(0) = i\}$. If $\mathbb{P}_i[T_i < \infty] = 1$ then the process is said to be *absorbing*.

In what follows, S will always be finite or countably infinite, and all measures will be defined on the discrete σ-algebra on $S \cup \{0\}$. As such, we will often identify the measure $\mathbf{u}$ with its associated probability vector and use the notation $\mathbf{u} = (u_i)_{i \in S}$ where $u_i := \mathbf{u}(\{i\})$ is the measure of the point i.

Definition 2.1.2. A Markov process $X = (X(t))_{t \geq 0}$ evolving on a space S is said to be *irreducible* if for every $i, j \in S$ there exists a sequence of states $s_1, \ldots, s_n \in S$ such that $q_{is_1} q_{s_1 s_2} \cdots q_{s_n j} > 0$, so that it is possible for the process starting from i to reach j in a finite amount of time. X is *reducible* otherwise. (As defined in Norris [1997].) We will often refer to S as being irreducible; S is irreducible if and only if X is.

Reducibility is equivalent to having that, for each $i, j \in S$ there exists $t \geq 0$ such that $\mathbb{P}_i[X(t) = j] > 0$. Note that this equivalence holds even if we consider a countably infinite state space. We now define the quasi-stationary distribution as described in

Collet et al. [2001].

Definition 2.1.3. Let $X = (X(t))_{t\geq 0}$ be a Markov process (discrete- or continuous-time) on a state space $\bar{S} = S \cup \{0\}$ where 0 is an absorbing state. Let $\mathbf{u}$ be a probability distribution on S. Then $\mathbf{u}$ is said to be a *Quasi-Stationary Distribution* (QSD) for X if for any $t \geq 0$ and any $\mathbf{u}$-measurable set A

$$\mathbb{P}_{\mathbf{u}}[X(t) \in A | X(t) \in S] = \mathbf{u}(A) \tag{2.1.3}$$

In a similar fashion to stationary distributions being linked to the limiting behaviour of a Markov process, one can consider the limiting behaviour of a transient Markov process conditional on the event of non-absorption.

Definition 2.1.4. Let $X = (X(t))_{t\geq 0}$ be a Markov process on a state space $\bar{S} = S \cup \{0\}$ where 0 is an absorbing state. Let $\mathbf{u}$ and $\mathbf{v}$ be probability distributions on S, then $\mathbf{u}$ is said to be the $\mathbf{v}$-*Limiting Conditional Distribution* ($\mathbf{v}$-LCD) for X if for any $\mathbf{u}$- and $\mathbf{v}$- measurable set A

$$\mathbf{u}(A) = \lim_{t\to\infty} \mathbb{P}_{\mathbf{v}}[X(t) \in A | X(t) \in S] \tag{2.1.4}$$

If $X(0) = i$ for some $i \in S$, then we may call it the i-LCD.

These LCDs are also referred to as *Yaglom limits*, as in Vere-Jones [1969]. Here we also briefly mention further properties related to quasi-stationarity, as described in van Doorn and Pollett [2013]. Firstly, *x-invariance* will be used to link the idea of quasi-stationarity to the eigenvector problems associated to the infinitesimal rate matrix $\bar{Q}$ and the transition probability matrix $P(t)$. For stationary distributions, one typically looks for solutions to the equation $\pi Q = 0$. In the case of quasi-stationarity, the conditioning on non-absorption leads us to look for eigenvectors for $\bar{Q}$ restricted to the transient states. Secondly, we introduce the *decay parameter* α, and the absorption parameter α_0 as defined in van Doorn and Pollett [2013]. We then define *a*-positive and *a*-recurrent processes.

Definition 2.1.5. Let $(X(t))_{t\geq 0}$ be a Markov process with infinitesimal rate matrix $\bar{Q}$ and transition probabilities $(P_{ij}(t))_{i,j\in\bar{S}}$ on a state space $\bar{S} = S \cup \{0\}$ with absorbing state 0. Moreover, let $x > 0$. Then a distribution $\mathbf{u} = (u_i)_{i\in S}$ is said to be *x-invariant for* Q if for each $j \in S$

$$\sum_{i\in S} u_i q_{ij} = -x u_j$$

A distribution $\mathbf{u}$ is said to be *x-invariant for* P for some $x > 0$ if, for each $j \in S$ and $t \geq 0$

$$\sum_{i \in S} u_i P_{ij}(t) = e^{-xt} u_j$$

If a distribution $\mathbf{u}$ is x-invariant for Q or for P, then x is said to be the *invariance value* of $\mathbf{u}$ for X.

Theorems 1 and 6 in van Doorn and Pollett [2013] show that these are equivalent on finite or irreducible state spaces.

Definition 2.1.6. Let $X(t)$ be an absorbing Markov process on state space $S \cup \{0\}$ with infinitesimal rate matrix $\bar{Q}$, where S is transient. If S is finite, then the *decay parameter* is $\alpha > 0$ where $-\alpha$ is the maximal non-zero eigenvalue of Q. If S is countable and irreducible, the decay parameter α is given by

$$\alpha := \inf\{a \geq 0 : P_{ij}(t) = O(e^{-at}) \text{ as } t \to \infty, \text{ for all } i, j \in S\} \tag{2.1.5}$$

We also use the alternative definition as seen in Anderson [1991]:

$$\alpha := -\lim_{t \to \infty} \frac{1}{t} \log P_{ij}(t) \geq 0 \tag{2.1.6}$$

This limit is independent of the choice of $i, j \in S$; see Kingman [1963].

The decay parameter, as can be seen from the definition, is the exponential rate of decay of the transition probabilities within S as mentioned in Kingman [1963]. According to Artalejo [2012], the decay parameter is also the exponential rate of extinction starting from the QSD on a finite state space. The decay parameter is often intractable when S is infinite, but some results have been found for specific birth-death processes which will be discussed in Section 3.1. In the finite case, it is theoretically computable as minus the least negative non-zero eigenvalue, but in practice solving eigenvalue problems for very large matrices is computationally expensive.

Definition 2.1.7. Let $(X(t))_{t \geq 0}$ be an absorbing Markov process on $S \cup \{0\}$, where S is irreducible with absorption time T. Then α_0, the *absorption parameter*, is given by

$$\alpha_0 = \inf\{a \geq 0 : \int_0^\infty e^{at} \mathbb{P}_i[T > t] dt = \infty\} \tag{2.1.7}$$

This is independent of the value of $i \in S$; see van Doorn and Pollett [2013].

The absorption parameter α_0 describes the rate at which the process converges to the absorbing state 0, by considering the distribution of the absorption time T. If S is finite, then $\alpha = \alpha_0$. This follows from Theorem 3.3.2 of Jacka and Roberts [1995] noting that since S is finite, the number of states from which S can be absorbed is finite. In the countably infinite case, we always have $\alpha \geq \alpha_0$ since

$$\alpha_0 = \liminf_t -t^{-1} \log \mathbb{P}_i[T > t] \leq \limsup_t -t^{-1} \log \mathbb{P}_i[T > t] \leq \alpha \tag{2.1.8}$$

because $\mathbb{P}_i[T > t] = \sum_{j \in S} P_{ij}(t) \geq P_{ij'}(t)$ for all $i, j' \in S$.

Definition 2.1.8. Let $(X(t))_{t \geq 0}$ be an absorbing Markov process on $S \cup \{0\}$ where S is irreducible. Then X is *a-positive* if for all $i \in S$

$$\lim_{t \to \infty} e^{at} P_{ii}(t) = 0.$$

X is said to be *a-recurrent* if, for all $i \in S$

$$\int_0^\infty e^{at} P_{ii}(t) = \infty.$$

Theorem 6 in van Doorn and Pollett [2013] links the ideas defined in this section.

Theorem 2.1.9. *Let $X = (X(t))_{t \geq 0}$ be an absorbing Markov process on $S \cup \{0\}$ with decay parameter $\alpha > 0$. A distribution $\mathbf{u}$ is a QSD for X if and only if it is a x-invariant distribution for P for some $x > 0$.*

Furthermore, $\mathbf{u}$ is a QSD for X if and only if $\mathbf{u}$ is x-invariant for Q for some $x > 0$. In this case, $x = \sum_{i=1}^\infty q_{i0} u_i$.

2.1.1 Existence and Uniqueness

In this section, we outline some preliminary results in the case of a finite or countable state space regarding existence and uniqueness for quasi-stationary distributions for certain classes of Markov processes. Firstly we consider the case of a finite, irreducible state space and state a result from Darroch and Seneta [1967].

Theorem 2.1.10. *Let $X = (X(t))_{t \geq 0}$ be an absorbing continuous time Markov process on finite state space $\bar{S} = S \cup \{0\}$ where $n = |S|$ and S is irreducible. Suppose*

X has infinitesimal rate matrix $\bar{Q}$. Then X has a unique QSD $\mathbf{u}$ on S. $\mathbf{u}$ is the unique positive solution:

$$\mathbf{u}^T Q = -\alpha \mathbf{u}^T \text{ with } \sum_{i=1}^{n} u_i = 1 \tag{2.1.9}$$

where α is the decay parameter of X and Q is the Q-matrix restricted to S. Furthermore, for any proper initial probability distribution $\mathbf{v}$ on S, the $\mathbf{v}$-LCD is equal to this QSD.

Reducible Finite State Spaces

In the case of a reducible finite state space, we follow the notation and results of van Doorn and Pollett [2008] and write the transient states S as a disjoint union of communicating classes $S = \bigcup_{l=1}^{L} S_l$. On these classes we put a partial ordering. We say $S_l \prec S_k$ if any state in S_k is accessible from any state in S_l (the process starting from S_k can enter S_l in finite time), and order the classes such that

$$S_l \prec S_k \Rightarrow l < k \tag{2.1.10}$$

This forces the Q-matrix to be block-lower-triangular.

For each communicating class S_l, we define the *decay parameter on leaving class* S_l by α_l, where $-\alpha_l$ is the maximal non-zero eigenvalue of $\bar{Q}$ restricted to S_l; this treats all other states as absorbing. Due to this block-diagonal structure, we must have $\alpha = \min_l \alpha_l > 0$.

Since two classes may have the same decay parameter, we denote by $S_{\min}$ the class with the lowest index l for which $\alpha_l = \alpha$. A class S_k is said to be *minimal for* α if it is minimal in the set $\{S_j : \alpha_j = \alpha\}$ with respect to the partial ordering $\prec$. We have that S_k is minimal for α if for all $l \neq k$ it is true that $S_l \prec S_k$ implies $\alpha_l > \alpha$. This minimality is not necessarily unique.

It is shown that if $\mathbf{u}$ is a QSD from which $S_{\min}$ is accessible (it is accessible from any state i for which $u_i > 0$), then it must satisfy $\mathbf{u}^T Q = -\alpha \mathbf{u}$ and $u_i > 0$ if and only if i is accessible from $S_{\min}$.

Theorem 2.1.11 (Theorem 5 van Doorn and Pollett [2008])**.** *Let $(X(t))_{t \geq 0}$ be an absorbing process on finite state space $\bar{S} = S \cup \{0\}$ where $S = \bigcup_{l=1}^{L} S_l$ with S_l as*

defined above and decay parameter $\alpha > 0$. Then $S_{\min}$ is the only minimal class for α if and only if there exists a unique QSD $\mathbf{u}$ on S from which $S_{\min}$ is accessible. In this case $\mathbf{u}$ can be renormalised to have unit mass. For this distribution, $u_i > 0$ if and only if $\{i\}$ is accessible from $S_{\min}$.

Moreover, suppose that α has geometric multiplicity one. Then for any initial distribution $\mathbf{v}$ from which $S_{\min}$ is accessible, the $\mathbf{v}$-LCD is precisely $\mathbf{u}$.

Example 2.1.12. Consider the process with state space $\{0, 1, 2, 3, 4\}$ and infinitesimal rate matrix $\bar{Q}$ in Figure 2.1.1. We see that each transient state is a communicating class, and that the decay parameter for this process is $\alpha = 1$. Although $S_{\min} = S_1 = \{1\}$, both S_1 and $S_2 = \{2\}$ are minimal for α. In this case we have non-uniqueness of the QSD, since we could, for example, put full mass at either one of the two states 1 or 2.

$$\bar{Q} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 & 0 \\ 1 & 0 & -1 & 0 & 0 \\ 0 & 2 & 0 & -2 & 0 \\ 0 & 0 & 2 & 0 & -2 \end{pmatrix} \begin{matrix} 0 \\ 1 \\ 2 \\ 3 \\ 4 \end{matrix}$$

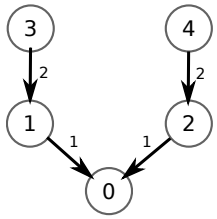


Figure 2.1.1: Infinitesimal Rate Matrix and Network for Example 2.1.12

Countable State Spaces

To extend our results to the countable state space, we return to the irreducible setting with a single, countably infinite communicating class $S = \{1, 2, \dots\}$. Note that in this case that even if 0 is absorbing, it is possible that absorption will not actually occur, and instead the process will move off to infinity; we set $T = \infty$ if absorption does not occur.

Using the decay parameter α as defined in (2.1.6), Theorem 4 of Vere-Jones [1969] links the ideas of x-invariance and quasi-stationarity in the infinite state space in the following theorem.

Theorem 2.1.13. *Let $X = (X(t))_{t \geq 0}$ be an absorbing Markov process on $S \cup \{0\}$ with decay parameter $\alpha > 0$. Let $\mathbf{u}$ be a QSD for X. Then $\mathbf{u}$ is x-invariant for Q*

for some $0 < x \leq \alpha$*, and in this case we have that*

$$x = \sum_{i \in S} u_i q_{i0} =: a_{\mathbf{u}} \tag{2.1.11}$$

Moreover, $\mathbf{u}$ *is* x*-invariant for* P*.*

However, this is only true in the case where $\alpha > 0$. In general, this may not be the case. Typically, we cannot evaluate α for such processes but we mention Theorem 3.3.2 of Jacka and Roberts [1995] which if the conditions are fulfilled implies $\alpha = \alpha_0$, a quantity that is often easier to calculate.

Theorem 2.1.14. *Let* $X = (X(t))_{t \geq 0}$ *be an absorbing Markov process on countable, irreducible* $S \cup \{0\}$ *with decay parameter* α*. Suppose that absorption is certain and the set of states* $\{s \in S : q_{s0} > 0\}$ *from which absorption can occur is finite. Then* $\alpha = \alpha_0$*.*

Unlike the finite case, existence of quasi-stationary distributions does not hold in general for a countable state space. Just within the class of birth-death processes, there exist examples of processes for which there exist no QSDs, a single unique QSD, or an uncountable collection of QSDs. In the case where S is irreducible, there are some known necessary and sufficient conditions for the existence of QSDs. To this end we use the following definition.

Definition 2.1.15. An absorbing Markov process $X = (X(t))_{t \geq 0}$ on $\mathbb{N} \cup \{0\}$ with absorption time T is said to satisfy *asymptotic remoteness* if for every $t \geq 0$ we have:

$$\lim_{i \to \infty} \mathbb{P}_i[T \leq t] = 0 \tag{2.1.12}$$

The following from van Doorn and Pollett [2013] outline some necessary conditions for the existence of QSDs on a countable, irreducible state space.

Theorem 2.1.16. *Let* $(X(t))_{t \geq 0}$ *be an absorbing Markov process on a countable irreducible state space* $S \cup \{0\}$ *with decay parameter* α*.*

- *Suppose there exists a QSD* $\mathbf{u}$ *for* X*. Then absorption is certain, and* $0 < a_{\mathbf{u}} \leq \alpha_0 \leq \alpha$*.*
- *If* $\mathbf{u}$ *is an* α*-invariant QSD for* X*, then* $\alpha_0 = \alpha = a_{\mathbf{u}}$ *as defined in* (2.1.11)*.*

- *Suppose* $\mathbf{u}$ *is a QSD, absorption is certain and asymptotic remoteness prevails. Then there exists* $i \in S$*, and* $\theta > 0$ *such that* $\mathbb{E}_i[e^{\theta T}] < \infty$*.*

Following the argument from Pollett [1999], we illustrate the necessary condition that a process must have certain extinction for a QSD to exist. Indeed, by Theorem 2.1.9, we know that since any QSD $\mathbf{u}$ is x-invariant for P for some $x > 0$, we also have, for $j \in S$, $t > 0$

$$\sum_{i \in S} u_i P_{ij}(t) = e^{-xt} u_j$$

Let $\sigma_i = \lim_{t\to\infty} P_{i0}(t)$ be the probability of extinction starting at i, then we know

$$\lim_{t\to\infty} \sum_{j \in S} P_{ij}(t) = 1 - \sigma_i$$

and so we have

$$0 \leq \sum_{j \in S} P_{ij}(t) \leq \frac{1}{u_i} e^{-xt} \sum_{j \in S} u_j = \frac{1}{u_i} e^{-xt} \tag{2.1.13}$$

since because S is irreducible, we must have $u_i > 0$ for all $i \in S$. Letting $t \to \infty$ we have $\sum_{j\in S} P_{ij}(t) \to 0$ and so $\sigma_i = 1$ for all i; in other words, extinction is certain.

Theorem 2.1.17. *Let* $X = (X(t))_{t\geq 0}$ *be an absorbing Markov process on countable irreducible state space* $S \cup \{0\}$ *with infinitesimal rate matrix* $\bar{Q}$ *and decay parameter* α*. Then the following hold.*

- *Suppose that absorption is certain,* $\alpha_0 > 0$ *and asymptotic remoteness holds as in Definition 2.1.15. Then there exists a QSD for* X*. Furthermore, suppose* X *has bounded waiting times* $-q_{ii} \leq C$ *for some* $C < \infty$ *independent of* $i \in S$*. Then* $\alpha_0 = \alpha$ *and there exists a QSD* $\mathbf{u}$ *which is* α*-invariant for* Q*.*

- *Suppose* X *is* α*-recurrent and that the number of states from which* 0 *is accessible is finite. Also suppose that absorption is certain and that* $\alpha > 0$*. Then* $\alpha_0 = \alpha$ *and there exists a unique QSD which is* α*-invariant for* Q*.*

- *Suppose that absorption is certain, asymptotic remoteness holds, and there exists* $i \in S$*,* $\theta > 0$ *such that* $\mathbb{E}_i[e^{\theta T}] < \infty$*. Then there exists a QSD for* X*.*

We conclude here by noting that, in general, these hold only for irreducible infinite

state spaces. There are not general results for absorbing processes on reducible countable state spaces.

2.1.2 Discrete-Time Processes

Although it is not the focus of this thesis, it is worth mentioning how the above concepts are different in the case of a discrete time process $(Y_t)_{t=0}^{\infty}$ with transition matrix $\bar{P} = [p_{ij}]_{i,j\in\bar{S}}$. In the finite case we can obtain, by Perron Frobenius theory, a maximal eigenvalue $\rho < 1$ for the sub-stochastic matrix $P = [p_{ij}]_{i,j\in S}$, the transition matrix restricted to S. Assuming S is irreducible and finite, there exist a unique left eigenvector $\mathbf{u}$ and right eigenvector $\mathbf{v}$ such that $\sum_i u_i = 1$ and $\sum_i u_i v_i = 1$. If Y is aperiodic, then we have

$$\lim_{t\to\infty} \rho^{-t} p_{ij}^t = v_i u_j$$

where $P^t = [p_{ij}^t]_{i,j\in S}$ is the t-step transition matrix. Given this, Section 3 of van Doorn and Pollett [1999] gives the following results:

Theorem 2.1.18. *Let $(Y_t)_{t\in\mathbb{N}}$ be an absorbing aperiodic Markov chain on a finite state space $\bar{S} = S \cup \{0\}$ with S irreducible. Then there exists a unique QSD $\mathbf{u}$ which is ρ-invariant for P in that for all $j \in S$, $t \geq 0$*

$$\mathbb{P}_{\mathbf{u}}[Y_t = j] = \rho^t u_j \qquad \sum_{i\in S} u_i p_{ij} = \rho u_j$$

Moreover, this QSD is the unique $\mathbf{v}$-LCD for the process from any proper initial condition $\mathbf{v}$.

Reducible State Spaces

For reducible state spaces, we again use the partial ordering $S_k \prec S_l$ as in (2.1.10). We denote by ρ_k the value of ρ for the process restricted to S_k. It can be shown that $\rho = \max\{\rho_l : 1 \leq l \leq L\}$ in this case. A class is said to be *maximal* if $\rho_k > \rho_l$ for all $l \neq k$ such that $S_l \prec S_k$. This leads to the following result from Theorems 4.1, 4.3 of van Doorn and Pollett [1999].

Theorem 2.1.19. *Let $(Y_t)_{t=0}^{\infty}$ be a Markov chain on a finite, reducible state space $S\cup\{0\}$ with $S = \bigcup_{l=1}^{L} S_l$ with partial ordering as in* (2.1.10)*. Then, for each maximal*

class S_k there exists a unique QSD $\mathbf{u}$ *for* Y *and* $\mathbf{u}$ *precisely gives weight to those states accessible from* S_k*.*

Moreover, let $k^* = \min\{k : \rho_k = \rho\}$ *and suppose* ρ *has geometric multiplicity one and* S_{k^*} *is aperiodic. Then for any initial distribution* $\mathbf{v}$ *from which* S_{k^*} *is accessible, the* $\mathbf{v}$*-LCDs exist and are all equal.*

2.2 Epidemic Models

In this section we discuss some basic properties of the density-dependent stochastic compartmental Susceptible-Infective-Removed epidemic models, including the construction, the basic reproductive number, and the distribution of the time to extinction for the SIS and SIR models. In particular, we make reference to the results and constructions of Andersson and Britton [2000], Kryscio and Lefèvre [1989] and Nåsell [2002].

2.2.1 SIS Epidemic Model

Consider a closed, homogeneously mixing population of N individuals, some of which are initially infected with a pathogen which is passed through contact, the rest initially susceptible to the infection. We denote the number of susceptibles at time t by $S(t)$ and the number of infectives by $I(t)$. Since the population is closed, $S(t) + I(t) = N$ for all $t \geq 0$ and due to this condition the process is entirely characterised by the behaviour of $I(t)$. We assume pairs of individuals make infectious contacts according to points of a homogeneous Poisson Process on $(0, \infty)$ with rate $\beta/N \geq 0$ for some $\beta > 0$. If a susceptible makes contact with an infective, they become immediately infected. Infections last for a random infectious period $L_I \sim \text{Exp}(\gamma)$ for some $\gamma > 0$, after which infectives become immediately re-susceptible. We assume complete independence of the contact processes and infection lifetimes. In summary, the epidemic $\mathbf{X}(t) = (I(t), S(t))$ proceeds according to the following events:

Infection Event:

$$\mathbb{P}[\mathbf{X}(t + \Delta t) = (i + 1, s - 1) | \mathbf{X}(t) = (i, s)] = \frac{\beta si}{N} \Delta t + o(\Delta t)$$

Recovery Event:

$$\mathbb{P}[\mathbf{X}(t + \Delta t) = (i - 1, s + 1) | \mathbf{X}(t) = (i, s)] = \gamma i \Delta t + o(\Delta t)$$

No event in $(t, t + \Delta t)$:

$$\mathbb{P}[\mathbf{X}(t + \Delta t) = (i, s) | \mathbf{X}(t) = (i, s)] = \left(1 - \left(\frac{\beta si}{N} + \gamma i\right) \Delta t\right) + o(\Delta t)$$

SIR Epidemic Model

For the SIR model, the process evolves as above but we instead assume that, following an infectious period, infectives are removed from the population, and never become re-susceptible; removal typically symbolising permanent immunity or death of the individual. We denote removed individuals by $R(t)$, and fix $S(t)+I(t)+R(t) = N$ for all $t \geq 0$. Contacts and infectious periods occur as for the SIS model, again with exponential waiting times. In this case, under the fixed population condition, the process is entirely characterised by the behaviour of $(I(t), S(t))$.

SIRS Epidemic Model

In this model we assume that, following an infectious period, the individual is temporarily removed, symbolising transient immunity to re-infection. This immunity lasts for an exponentially distributed length of time $L_R \sim \text{Exp}(\delta)$ for some $\delta > 0$, and all immune periods are all independent from each other and from all other behaviour of the model. We denote the number of temporarily immune individuals by $R(t)$, and observe that the process is entirely characterised by $(I(t), S(t))$. Under these interactions, the epidemic $\mathbf{X}(t) = (I(t), S(t), R(t))$ proceeds according to the following events.

Infection Event:

$$\mathbb{P}[\mathbf{X}(t+\Delta t) = (i+1, s-1) | \mathbf{X}(t) = (i,s)] = \frac{\beta si}{N}\Delta t + o(\Delta t)$$

Recovery Event:

$$\mathbb{P}[\mathbf{X}(t+\Delta t) = (i-1, s) | \mathbf{X}(t) = (i,s)] = \gamma i \Delta t + o(\Delta t)$$

Loss of Immunity Event:

$$\mathbb{P}[\mathbf{X}(t+\Delta t) = (i, s+1) | \mathbf{X}(t) = (i,s)] = \delta(N-(i+s))\Delta t + o(\Delta t)$$

No event in $(t, t+\Delta t)$:

$$\mathbb{P}[\mathbf{X}(t+\Delta t) = (i,s) | \mathbf{X}(t) = (i,s)] = 1 - \left(\frac{\beta si}{N} + \gamma i + \delta(N-(i+s))\right)\Delta t + o(\Delta t)$$

2.2.2 QSDs for Standard Epidemic Models

In this section, we briefly discuss the existence and uniqueness of QSDs for the SIS, SIR and SIRS epidemic models.

Proposition 2.2.1. *For the SIS epidemic model $I(t)$ as defined in Section 2.2.1 with β, $\gamma > 0$, $N \in \mathbb{N}$ there exists, conditional on $\{I > 0\}$, a unique QSD which gives mass to all states in $\{1, \dots, N\}$.*

This follows directly from Theorem 2.1.10.

Proposition 2.2.2. *For the SIR epidemic model $(I(t), S(t))$ as defined in Section 2.2.1 with β, $\gamma > 0$ and $N \in \mathbb{N}$ there exists, conditional on $\{I > 0\}$, a unique QSD which gives full weight to the state $(1, 0)$ with a single infective and no susceptibles.*

This follows directly from Theorem 2.1.11 noting that each state (i, s) is its own communicating class with decay parameter $\beta si/N + \gamma i$ which is minimised with $s = 0, i = 1$. This state can only reach the absorbing state, so must have full weight.

Proposition 2.2.3. *For the SIRS epidemic model $(I(t), S(t))_{t \geq 0}$ with β, γ, $\delta > 0$ and $N \in \mathbb{N}$, we consider two cases. Conditional on $\{I > 0\}$, there exists a unique QSD $\mathbf{u}_A$ which gives weight to all states $\{(i, s) : i > 0\}$. Conditional on $\{S < N\}$ there exists a unique QSD $\mathbf{u}_B$. If α_A is the decay parameter associated to $\mathbf{u}_A$, then $\mathbf{u}_B$ gives full weight to all states only if $\alpha_A < \delta$. Otherwise, full weight goes to the state $(0, N - 1)$.*

This again follows directly from Theorem 2.1.11.

2.2.3 The Basic Reproductive Number

In all of these models, there are two main behaviour regimes, determined by a threshold parameter R_0, known as the *basic reproduction number*, defined to be the expected number of secondary infections occurring during a typical individual's infectious period in an otherwise susceptible infinite population. For the three above models we have $R_0 = \beta/\gamma$. If $R_0 < 1$ then one expects an epidemic starting from a single infective to die out quickly only ever reaching a small proportion of the population. Conversely, if $R_0 > 1$ then there is a non-zero probability that the

epidemic will affect a large proportion of the population (of the order of the population size). To quantify the probability of a large epidemic, one can use branching process approximations as in Andersson and Britton [2000]. These approximating processes have, in the case of $R_0 > 1$, a probability of $1 - (1/R_0)^{I(0)}$ of the number of infectives going to infinity as $t \to \infty$.

One way to discuss the epidemic quantitatively is to consider the time to extinction, typically from an epidemic starting from a single infective.

Proposition 2.2.4 (Theorem 8.1, Andersson and Britton [2000])**.** *For the SIS model defined in Section 2.2.1, let $R_0 = \beta/\gamma$, and let $T^{(N)}$ be the extinction time of the SIS model with a population of size N and initial condition $I(0) = 1$. Then the following hold:*

1. *If $R_0 > 1$, then $T^{(N)} \to T_0$ almost surely as $N \to \infty$ and $\mathbb{P}_1[T_0 < \infty] = \gamma/\beta < 1$. Here, T_0 is the extinction time for a Linear BDP as defined in 3.1 $Y(t)$ with $Y(0) = 1$, birth rate β and death rate γ.*
2. *If $R_0 \leq 1$, then $T^{(N)} \to T_0$ a.s. and $\mathbb{P}_1[T_0 < \infty] = 1$.*

Proposition 2.2.5 (Theorem 4.3, Andersson and Britton [2000])**.** *For the SIR model defined in Section 2.2.1, let $R_0 = \beta/\gamma$, and let $T^{(N)}$ be the extinction time of the SIR model with a population of size N and $(I(0), S(0)) = (1, N-1)$. Then the following hold:*

1. *If $R_0 \leq 1$, then $T^{(N)} \to T_0$ almost surely as $N \to \infty$, where T_0 is the extinction time for a Linear BDP $Y(t)$ with $Y(0) = 1$, birth rate β and death rate γ.*
2. *If $R_0 > 1$ then, on the event $\{T_0 < \infty\}$ which has probability $\frac{1}{R_0}$, $T^{(N)} \to T_0$ almost surely as above. On the event $\{T_0 = \infty\}$, we have*

$$T^{(N)} - \left(\frac{1}{\gamma - \beta(1-\tau)} + \frac{1}{\beta - \gamma}\right)\log(N) - C_{\beta,\gamma} \to W$$

where τ is the non-zero solution to $1 - e^{-\gamma\beta\tau} = \tau$, $C_{\beta,\gamma}$ is a constant, and W is a random variable defined by

$$W = \frac{W_1}{\gamma - \beta(1-\tau)} + \frac{W_2}{\gamma - \beta}$$

where W_1, W_2 are independent and have the same cumulative distribution $F(x) = \exp(e^{-x})$ for $x > 0$.

2.3 Sequential Monte Carlo

Stochastic epidemiology frequently presents intractable probability densities, such as those in the SIRS epidemic model, or intractable likelihoods in the case of more complex models with heterogeneity between individuals, as described in Nåsell [2002]. In order to draw from complicated distributions, or those we don't have an analytic expression for, we make use of probabilistic Monte Carlo methods. In particular, we make use of concepts related to Sequential Monte Carlo, a concept first mentioned by Gordon et al. [1993], and investigated by Doucet et al. [2001] and Liu and Chen [1998]. In this section we outline the basic idea of Sequential Monte Carlo (SMC), also known as *Particle Filtering*. Throughout, we assume that all probability measures of interest p have a density given by $p(x)$. As in Section 2.1, on countable spaces S we identify the measure p with the probability vector $p = (p(x))_{x \in S}$.

A typical setting for standard SMC methods is that of the Hidden Markov Model. In this setting, as described in Doucet et al. [2001], a Markov chain $X = (X_n)_{n \geq 0}$ with initial distribution $f_X(x_0)$ evolves over time according to a given density $f_X(x_n|x_{n-1})$. However, we only observe some sequence $Y = (Y_n)_{n \geq 0}$. Here we assume the observations are independent conditional on the X-process and for each n the observation is distributed according to probability density $f_Y(y_n|x_n)$. Using the notation

$$\mathbf{x}_{n_1:n_2} = (x_{n_1}, x_{n_1+1}, \ldots, x_{n_2-1}, x_{n_2}) \qquad 0 \leq n_1 \leq n_2$$

one is typically interested in evaluating a *target distribution* $\pi_n(\mathbf{x}_{1:n}|\mathbf{y}_{1:n})$, the distribution of the X-process given a sequence of observations $\mathbf{Y}_{1:n} = \mathbf{y}_{1:n}$. Given this setup it is possible to obtain iterative formulae for $\pi(\mathbf{x}_{1:n}|\mathbf{y}_{1:n})$, making use of Bayes' rule:

$$\pi(\mathbf{x}_{1:n}|\mathbf{y}_{1:n}) \propto p(\mathbf{y}_{1:n}|\mathbf{x}_{1:n})p(\mathbf{x}_{1:n})$$

This can then be normalised and iterated backwards in time to give

$$\pi(\mathbf{x}_{1:n}|\mathbf{y}_{1:n}) = \pi(\mathbf{x}_{1:n-1}|\mathbf{y}_{1:n-1}) \frac{p(y_n|x_n)p(x_n|x_{n-1})}{p(y_n|\mathbf{y}_{1:n-1})}.$$

Often, one can't evaluate $\pi(\mathbf{x}_{1:n}|\mathbf{y}_{1:n})$ directly, so we employ Monte Carlo methods such as Importance Sampling to obtain approximate draws from the target distri-

bution.

2.3.1 Importance Sampling

Following the outline in Doucet and Johansen [2009], suppose that one wishes to obtain $\mathbb{E}[h(X)]$ where $X \sim \pi$, some *target distribution* defined on a measurable space $(E, \mathcal{F})$. Further suppose $\pi(x)$ is only known up to some normalisation constant Z; we denote the unnormalised density by $\gamma(x) := \pi(x)Z$. In importance sampling, we make use of a *proposal distribution* $q(x)$ which is typically known and easier to draw from than π. Assuming that the support of the target distribution is entirely contained within the support of the proposal, one can obtain the identity

$$\mathbb{E}[h(X)] = \frac{\int h(x)w(x)q(x)dx}{\int w(x)q(x)dx}$$

where $w(x)$ is the *importance weight* function given by:

$$w(x) = \frac{\gamma(x)}{q(x)} \tag{2.3.1}$$

A number of samples $\{x^i : i = 1, \dots M\}$ are drawn according to q and then we obtain a Monte Carlo estimate for the target distribution:

$$\widehat{\pi^M}(x) = \sum_{i=1}^{M} W^i \delta_{x^i}(x)$$

where $W^i = w(x^i)/\sum_{j=1}^{M} w(x^j)$ are the normalised importance weights and $\delta_y(x)$ is the Dirac point measure at y. Integrating h with respect to $\widehat{\pi}$ leads us to an estimate $\hat{I}^M(h(X))$ for the expectation:

$$\hat{I}^M(h(X)) = \sum_{i=1}^{M} W^i h(x^i)$$

Kolmogorov's strong law of large numbers shows that, asymptotically, we have almost-sure convergence of $\hat{I}^M(h(X))$ to the true value as $M \to \infty$, as long as the true expectation $\mathbb{E}[h(X)]$ exists and is finite. In practice, a finite number of particles M leads to a biased estimate arising as the ratio of two estimates, but Geweke [1989] gives a central limit theorem which says that, as long as $\mathbb{E}[w(X)]$ and

$\mathbb{E}[h^2(X)w(X)]$ are finite, then

$$M^{1/2}\left(\hat{I}^M(h(X)) - \mathbb{E}[h(X)]\right) \xrightarrow{d} N(0, \sigma^2)$$

where σ^2 is the true variance of $h(X)$.

2.3.2 Sequential Importance Sampling

Following work from Doucet and Johansen [2009], one can extend importance sampling to draw instead from a sequence of target distributions $\{\pi_n : 1 \leq n \leq N\}$. Here, π_n is defined on measurable space $(E_n, \mathcal{F}_n)$, where $E_n = E^n$ is the Cartesian product of some space E, and $\mathcal{F}_n = \mathcal{F}_{n-1} \times \mathcal{F}$ is the n-fold product σ-algebra of $\mathcal{F}$ defined on E. We assume each density $\pi_n(\mathbf{x}_{1:n}) : E_n \to \mathbb{R}$ is known only up to some normalisation constant Z_n; we denote the unnormalised density by $\gamma_n(\mathbf{x}_{1:n})$ as in the previous section. Under importance sampling, we would select a sequence of proposal densities $\{q_{1:n}(\mathbf{x}_{1:n}) : 1 \leq n \leq N\}$ from which to draw, and compute at each step the importance weights defined on the whole product space. To reduce such computation costs, we make a specific choice of proposal:

$$q_n(\mathbf{x}_{1:n}) = q_1(x_1)\prod_{i=1}^{n} q_i(x_i|x_{i-1}) \tag{2.3.2}$$

From this, we see that to obtain draws from q_n, we can use the previous draws $\{X_{n-1}^{(i)} : 1 \leq i \leq M\}$ and draw from the conditional distribution $X_n^{(i)}|X(i)_{n-1} \sim q_n(x_n|X_{n-1}^{(i)})$ instead. Substituting (2.3.2) into the definition of the importance weight (2.3.1) gives

$$\begin{aligned}
w_n(\mathbf{x}_{1:n}) &= \frac{\gamma_n(\mathbf{x}_{1:n})}{q_1(x_1)\prod_{i=1}^{n} q_i(x_i|x_{i-1})} \\
&= \frac{\gamma_{n-1}(\mathbf{x}_{1:n-1})}{q_1(x_1)\prod_{i=1}^{n-1} q_i(x_i|x_{i-1})} \frac{\gamma_n(\mathbf{x}_{1:n})}{\gamma_{n-1}(\mathbf{x}_{1:n-1})q_n(x_n|x_{n-1})} \\
&= w_{n-1}(\mathbf{x}_{1:n-1})\frac{\gamma_n(\mathbf{x}_{1:n})}{\gamma_{n-1}(\mathbf{x}_{1:n-1})q_n(x_n|x_{n-1})} \\
&=: w_{n-1}(\mathbf{x}_{1:n-1})\widetilde{w}_n(x_n, \mathbf{x}_{1:n-1})
\end{aligned} \tag{2.3.3}$$

which gives us an iterative form for the importance weight, computing $\widetilde{w}_n(x_n, \mathbf{x}_{1:n-1})$, the *incremental weight* at each step. Given this, we can renormalise the importance

weights to get $(W_n^1, \ldots, W_n^M)$ and then obtain estimates for the target distributions:

$$\pi_n^{(M)}(\mathbf{x}_{1:n}) = \sum_{j=1}^{M} W_n^j \delta_{\mathbf{X}_{1:n}^j}(\mathbf{x}_{1:n})$$

Algorithm 2.3.1 Sequential Importance Sampling to draw from $\{\pi_n = \gamma_n/Z_n\}_{n=1}^N$

Require: $M \geq 1$, $q_1(x_1)$ probability measure, $q_k(x_k|x_{k-1})$ conditional probability measures for $k = 1, \ldots, N$,

1: Draw $X_1^j \sim q_1$ for $j = 1, \ldots M$.
2: For $j = 1, \ldots, M$ compute $w_1(X_1^j) = \gamma_1(X_1^j)/q_1(X_1^j)$ and normalise to give $W_1^j := w_1(X_1^j)/\sum_{j=1}^M w_j(X_1^j)$.
3: **for** $n = 2$ **to** N **do**
4: For $j = 1, \ldots, M$ draw $X_n^j \sim q_n(x_n|x_{n-1} = X_{n-1}^j)$.
5: For $j = 1, \ldots, M$ compute $\widetilde{w}_n(X_n^j, \mathbf{X}_{1:n-1}^j)$ from (2.3.3) and normalise to get W_n^j.
6: **end for**
7: **return** $\pi_n^{(M)}(x_{1:n}) = \sum_{j=1}^M W_n^j \delta_{\mathbf{X}_{1:n}^j}(x_{1:n})$ for $n = 2, \ldots, N$.

In the case of the Hidden Markov Model, we have proposal distribution $q_{1:n}(\mathbf{x}_{1:n}|\mathbf{y}_{1:n})$ which can be rewritten as

$$q_{1:n}(\mathbf{x}_{1:n}|\mathbf{y}_{1:n}) = q_n(x_n|\mathbf{x}_{1:n-1}, \mathbf{y}_{1:n}) q_{1:n-1}(\mathbf{x}_{1:n-1}|\mathbf{y}_{1:n-1})$$

and so the importance weights are given by

$$w_n(\mathbf{x}_{1:n}) = w_{n-1}(\mathbf{x}_{1:n}) \frac{f_Y(y_n|x_n) f_X(x_n|x_{n-1})}{q_n(x_n|\mathbf{x}_{1:n-1}, \mathbf{y}_{1:n})}$$

If one chooses the proposal distribution to be

$$q_{1:n}(\mathbf{x}_{1:n}|\mathbf{y}_{1:n}) = f_X(x_1) \prod_{i=2}^{n} f_X(x_i|x_{i-1})$$

in the case where we know how to draw from $f_X(x_n)$ directly, then the importance weight simplifies to $w_n(\mathbf{x}_{1:n}) \propto w_{n-1}(\mathbf{x}_{1:(n-1)}) f_Y(y_n|x_n)$. In terms of the limiting estimates of the expectation as $N \to \infty$, this choice minimises the variance, although is rarely possible to compute. More generally, as $N \to \infty$, Crisan and Doucet [2000] says that for a fixed number of particles M:

$$\mathbb{E}\left[(\hat{I}^M(h(\mathbf{X}_{1:N})) - \mathbb{E}[h(\mathbf{X}_{1:n})])^2\right] \leq C_N M^{-1} \|h\|_2^2$$

One should note however, that C_N could grow exponentially in N, and so equivalent precision requires many more particles as N increases.

2.3.3 Particle Resampling

Standard SMC methods, such as Sequential Importance Sampling, suffer from what is referred to by Collet et al. [2001] as *particle weight degeneracy* where, for long time series (i.e. for large N in the previous section), one will end up with a single particle with nearly all the weight (after normalisation), and many extremely low-weighted particles.

One method of dealing with the problem of particle weight degeneracy is through *Particle Resampling*, also known as bootstrapping. To implement a particle resampling scheme, we define a sequence of resampling timepoints $\{\tau_1, \tau_2, \ldots, \tau_k\}$. At each resampling timepoint, we redraw M' particles with replacement from a pool of M particles $\{(X_j, W_j) : j = 1, \ldots, M\}$ with normalised weights.

The sequence of resampling timepoints can be determined either deterministically or dynamically according to some given distribution. One of the most common dynamic implementations is through the use of a *threshold parameter*. At each of a deterministic sequence of timepoints, a given statistic is computed, and resampling occurs if the statistic falls outside some predefined region. Typically, this is implemented by resampling whenever the effective sample size falls below some proportion of the total sample size. Other common statistics, include the *coefficient of variation*

$$cv^2 = \frac{\sum_{j=1}^{M}(W^j - \bar{W})^2}{\bar{W}^2}$$

where $\bar{W}$ is the mean of the importance weights, and the *relative entropy*

$$\text{Ent} = \sum_{j=1}^{M} -\log(W^j)/M$$

In most circumstances, at the resampling step we will assume $M' = M$, maintaining a constant number of particles. For the moment, however, we maintain full generality.

One desirable property in a resampling method is the preservation of properly weighted samples, as defined in Liu and Chen [1998].

Definition 2.3.1. A sample of weighted particles $\{(X_i, w_i) : 1 \leq i \leq M\}$ with normalised weights and with $X_i \sim f$ for all $1 \leq i \leq M$ is said to be *properly weighted* with respect to a measure π if, for any square integrable function h there exists $k > 0$ such that, if $Y \sim \pi$,:

$$\mathbb{E}\left[\sum_{i=1}^{M} w_i h(X_i)\right] = k\mathbb{E}[h(Y)]$$

Here we outline a few commonly used resampling methods that preserve proper weighting. They are demonstrated graphically in Figure 2.3.1, where we assume $M = M' = 8$ and consider a collection of weighted samples $\{(X^j, W^j) : j = 1, \ldots, 8\}$.

Definition 2.3.2 (Multinomial resampling.)**.** Taking the definition from Liu and Chen [1995], one can draw M' particles with replacement from the M current normalised weighted particles $\{(X_j, W_j) : 1 \leq j \leq M\}$ independently and proportionally according to their current importance weight. In particular, if N_j denotes the number of resampled copies of X_j, then $\mathbf{N} = (N_1, \ldots, N_M)$ is distributed according to the multinomial distribution $\mathbf{N} \sim \text{Multinomial}(M'; \{W_1, \ldots, W_M\})$.

Definition 2.3.3 (Stratified resampling.)**.** Taking the definition from Douc and Cappé [2005], we let, for $l = 1, \ldots, M'$, $\mathcal{U}_l \sim \text{Unif}(\frac{l-1}{M'}, \frac{l}{M'})$. For $j = 1, \ldots, M$, we then select N_j copies of particle X_j by choosing

$$N_j = \left| \left\{ l : \sum_{k=1}^{j-1} W_k < \mathcal{U}_l \leq \sum_{k=1}^{j} W_k \right\} \right|,$$

where an empty sum is defined to be equal to 0. This method ensures that $N_j - 1 \leq M'W_j = \mathbb{E}[N_j] \leq N_j + 1$ and at least one copy is retained of every particle with $W_j \geq \frac{2}{M'}$. This divides the importance weight into equal $1/M'$ regions, and draws one particle from each region.

Definition 2.3.4 (Systematic resampling.)**.** Systematic resampling works in the same way as Stratified except we draw $\mathcal{U}_1 \sim \text{Unif}(0, \frac{1}{M'})$ and then deterministically choose $\mathcal{U}_l = \mathcal{U}_1 + \frac{l-1}{M'}$ and proceed as before. This ensures that at least one copy is retained of every particle with $W_j \geq \frac{1}{M'}$.

Definition 2.3.5 (Residual resampling.)**.** Taking the definition from Liu and Chen [1998], we first let $N_{j,a} = \lfloor M'W_j \rfloor$ for $j = 1, \ldots, M$, where $\lfloor x \rfloor$ is the integer part of

x. Then we draw $\mathbf{N}_b = (N_{1,b}, \ldots, N_{M,b})$ from the multinomial distribution

$$\mathbf{N}_b \sim MN\left(M' - \sum_{k=1}^{M} \lfloor M' W_k \rfloor; \left\{ \frac{M' W_j - \lfloor M' W_j \rfloor}{\sum_{k=1}^{M} (M' W_k - \lfloor M' W_k \rfloor)} \right\}_{j=1}^{M}\right)$$

We then resample by taking $N_j = N_{j,a} + N_{j,b}$ copies of particle X_j. In practise, this assigns one copy of a particle for each $1/M$ of the total normalised weight it has, then the rest are drawn multinomially from the residual importance weight.

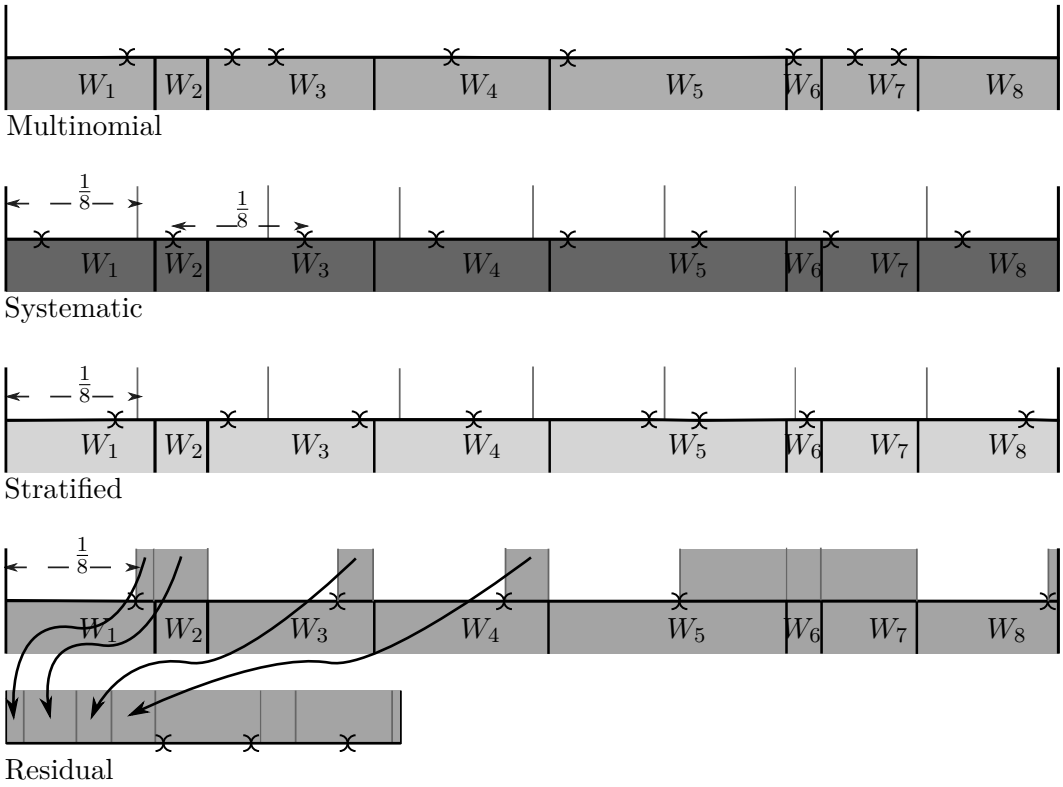


Figure 2.3.1: Comparison of Multinomial, Systematic, Stratified and Residual Resampling Schemes: Drawing 8 new particles from 8 weighted particles $\{(X_j, W_j) : j = 1, \ldots, 8\}$

It should be noted that with the above techniques, it is known that there are some guidelines one should follow when implementing resampling methods. Firstly, there is clearly nothing to be gained by resampling at the final step, and indeed, whenever a sample is take, it should be done so before resampling, not after. This comes from the fact that all the above resampling techniques increase Monte Carlo variance by potentially reducing the number of distinct particle histories through failure to select enough unique particles. This increases the variance of estimate for statistics such

as the sample mean. Hence, resampling too often, besides being potentially computationally expensive, can lead to poorer Monte Carlo estimates. On the other hand, resampling too infrequently will still cause some level of particle weight degeneracy between resampling steps. Since the resampled particles are typically drawn from the higher weight particles, this would lead again to an impoverished selection of particle histories. Liu and Chen [1995] explain this in more depth.

Chapter 3

Characterisation of QSDs for Birth-Death and Epidemic Models

This chapter concerns itself with the development of results regarding characterising the behaviours of QSDs in the case of birth-death processes (BDPs), and in particular Linear BDPs. A lot of the surrounding literature such as Collet et al. [2001] and Cavender [1978] focuses on the existence and uniqueness of such QSDs. Moreover, when characterisation is considered, the main focus is on bounds for the decay parameter associated to the process. Both Sirl et al. [2007] and van Doorn [1991] focus on this aspect with different approaches, related to the study of orthogonal polynomials associated to the process.

We will start the chapter by making reference to existing work and define the process, noting characteristics of its behaviour. We link this to the stochastic epidemic models we discussed in Section 2.2.1. We will discuss the known results for existence and uniqueness of QSDs for general BDPs as discussed by Collet et al. [2001] before extending current work by characterising the QSDs for the linear birth-death process. First we consider the QSD $\mathbf{u}(\alpha)$ which corresponds to $\mathbf{v}$-LCDs with finite support for $\mathbf{v}$ as discussed in Zhang and Liu [2012]. Then we will consider the collection $\{\mathbf{u}(x) : x < \alpha\}$, which correspond to $\mathbf{v}$-LCDs for distributions $\mathbf{v}$ with infinite mean. We give characterisations of these processes depending on the birth and death rates β and γ, and x.

Following this, there will be a more in-depth look at the general death processes. This will be prefaced by some existing results from van Doorn and Pollett [2008]. We will then characterise the $\mathbf{v}$-LCDs with initial distributions $\mathbf{v}$ with finite support, and mention some cases where we can say something about the x-invariant QSDs $\mathbf{u}(x)$ in this case. The linear death process will be discussed in Section 3.3.1 as a special case for which we can say a little more.

Finally, we will consider the SIS epidemic model defined in Section 2.2.1 as a birth-death process on a finite state space, and analyse two related recurrent processes whose stationary distributions approximate the QSD for the SIS model, similar to those in Clancy and Pollett [2003]. We will obtain exact expressions for these approximations and mention how the modes of the processes converge for population limits. Note that throughout this chapter, any pre-existing results are cited - all other results are my own.

3.1 Construction and Basic Results

According to Anderson [1991], a *Birth-Death Process* (BDP) $X = (X(t))_{t\geq 0}$ is a Markov Process on $\mathbb{N} \cup \{0\}$ characterised by the following events.

Infection Event:

$$\mathbb{P}[X(t+\Delta t) = i+1 | X(t) = i] = \beta_i \Delta t + o(\Delta t)$$

Recovery Event:

$$\mathbb{P}[X(t+\Delta t) = i-1 | X(t) = i] = \gamma_i \Delta t + o(\Delta t)$$

No event in $(t, t+\Delta t)$:

$$\mathbb{P}[\text{no event in}(t, t+\Delta t) | X(t) = i] = (1 - (\beta_i + \gamma_i)\Delta t) + o(\Delta t)$$

Condition of no multiple events: The probability of multiple events occurring in any time interval $(t, t+\Delta t]$ is $o(\Delta t)$.

for a given sequence of non-negative birth rates $(\beta_i)_{i\geq 1}$ and death rates $(\gamma_i)_{i\geq 1} \geq 0$.

We assume that birth-death processes are absorbing, so β_0 is fixed to be zero. When $\beta_i = i\beta$, $\gamma_i = i\gamma$ we will refer to this as the *Linear BDP*.

To motivate its inclusion in this thesis, we note that the Linear Birth-Death Process can arise in two particular processes of biological interest.

BGW Branching Process Firstly, we compare the Linear BDP to the *Bienaymé-Galton-Watson* (BGW) branching process, discussed in Athreya and Ney [1972]. In this process, each individual lives for an exponentially distributed lifetime $L \sim \text{Exp}(\lambda)$, and then gives birth to a number of offspring according to some distribution, characterised by a probability generating function $\mathcal{F}(s)$. However, if we set $\lambda = \beta + \gamma$ and $\mathcal{F}(s) = \frac{\gamma+\beta s^2}{\gamma+\beta}$, one obtains a branching process which behaves the same as the Linear BDP: a birth is equivalent to two offspring, and a death is equivalent to no offspring for a given individual. This is a natural process for discussing the growth of a population or infection.

SIS as Birth-Death Process We will also look at this process from a more epidemiologically relevant perspective. If we consider the SIS model as defined in Section 2.2 with N individuals, one starts with some fixed number of initial infectives $I(0) = i$. If one keeps this initial number of infectives i constant, then as $N \to \infty$ we get that $S(0)/N \to 1$ (recall that $I(0) + S(0) = N$). If we consider the rate of infections, we get $\beta I(t)S(t)/N \to \beta I(t)$ as $N \to \infty$, but the rate of recovery is independent of N and so stays constant at $\gamma I(t)$. We now see that these are precisely the rates for the linear birth-death process: births happening with rate $\beta I(t)$, deaths with rate $\gamma I(t)$.

Intuitively this can be seen from the idea that as the initial number of susceptibles grows, there becomes an infinite pool of susceptibles to make contact with. As a result, every infectious contact is successful since the probability of two infectives making contact is zero as $N \to \infty$. Consequently, infections occur from each infective with constant rate β, leading to our conclusion above. Conversely, one could consider the SIS model to be a special case of the general BDP with rates $\beta_i = \beta i(N-i)/N$ and $\gamma_i = i\gamma$, restricting the process to the states $\{0, \ldots, N\}$.

3.1.1 QSDs for General Birth Death Processes

Before we consider the Linear BDP more closely, it is worth noting some of the more general results regarding existence and uniqueness for QSDs of general BDPs.

Theorem 3.1.1 (Theorem 3.2 van Doorn [1991]). *Let $(X(t))_{t\geq 0}$ be a general birth death process with birth rates $(\beta_i)_{i\geq 1} > 0$ and death rates $(\gamma_i)_{i\geq 1} > 0$ and decay parameter α. Define the sequence $(\pi_n)_{n\geq 1}$ by*

$$\pi_1 = 1 \qquad\qquad \pi_n = \frac{\beta_1 \dots \beta_{n-1}}{\gamma_2 \dots \gamma_n}, \quad n > 1$$

and let

$$\mathcal{A} := \sum_{n=1}^{\infty} \frac{1}{\beta_n \pi_n} \sum_{i=n+1}^{\infty} \pi_i \tag{3.1.1}$$

Then the following hold:

- *Suppose that extinction is certain, and the series $\mathcal{A}$ converges. Then the decay parameter $\alpha > 0$ and there exists a unique quasi-stationary distribution $\mathbf{u}$.*
- *Suppose extinction is certain, the series $\mathcal{A}$ diverges and $\alpha = 0$. Then no quasi-stationary distribution exists.*
- *Suppose extinction is certain, the series $\mathcal{A}$ diverges and $\alpha > 0$. Then there exist a family of quasi-stationary distributions $\{\mathbf{u}(x) : 0 < x \leq \alpha\}$. Specifically, $\mathbf{u}(x)$ is x-invariant for Q, where for $j \in S$*

$$\sum_{i\in S} u_i q_{ij} = -x u_j \qquad \text{with } \sum_{i\in S} u_i = 1.$$

Here we note that, in general, there is no way to obtain α directly from the series $\mathcal{A}$. Some work has been done by van Doorn [2015] in estimating the decay parameter from the birth and death rates, but as mentioned before, α in general has no expression in terms of standard functions.

With regards to the limiting conditional distributions of the general birth-death process, only limited results are known. In the case above where the QSD is unique, it is clearly also the unique LCD. If the QSD is not unique, then van Doorn [1991] shows that the $\mathbf{v}$-LCD for any initial distribution $\mathbf{v}$ with finite support is the α-invariant QSD. In general, the domains of attraction for the different QSDs are not

known, but in the special cases of the Linear BDP and the general death process we will extend these results.

3.2 The Linear Birth-Death Process

In this section we focus our attention to the linear case of the birth-death process. We first explore the basic properties of the process.

One can obtain from the Kolmogorov forward equations the expected behaviour of the process which tells us that $\mathbb{E}[X(t)] = X(0)e^{(\beta-\gamma)t}$, suggesting exponential growth in the case where $\beta > \gamma$, and decay when $\beta < \gamma$. Furthermore, one can also obtain the generating function $\mathcal{G}(s,t) = \mathbb{E}[s^{X(t)}]$ for the Linear BDP.

Theorem 3.2.1 (Anderson [1991])**.** *The probability generating function for the linear BDP $X = (X(t))_{t\geq 0}$ with parameters $\gamma, \beta > 0$ and $\gamma \neq \beta$, and $X(0) = x$ is given by*

$$\mathcal{G}(s,t) = \mathbb{E}[s^{X(t)}] = \left(\frac{\gamma(s-1)e^{(\beta-\gamma)t} - (\beta s - \gamma)}{\beta(s-1)e^{(\beta-\gamma)t} - (\beta s - \gamma)}\right)^x \tag{3.2.1}$$

If $\gamma = \beta$ then

$$\mathcal{G}(s,t) = \left(\frac{\gamma t + s(1-\gamma t)}{1 + \gamma t - \gamma t s}\right)^x$$

Consequently, we have that the expected behaviour of the process is given by

$$\mathbb{E}_x[X(t)] = xe^{(\beta-\gamma)t} \tag{3.2.2}$$

for $t \geq 0$ and the probability of non-absorption before time $t \geq 0$ is given by

$$\mathbb{P}_x[T > t] = \begin{cases} 1 - \left(\frac{(\beta-\gamma)e^{(\beta-\gamma)t}}{\beta e^{(\beta-\gamma)t} - \gamma}\right)^x & \beta \neq \gamma \\ 1 - \left(\frac{\gamma t}{1+\gamma t}\right)^x & \beta = \gamma \end{cases} \tag{3.2.3}$$

The probability generating function follows from the Kolmogorov forward equations, which can be used to derive the differential equation

$$\frac{\partial}{\partial t}\mathcal{G}(s,t) = (\gamma - \beta s)(1-s)\frac{\partial}{\partial s}\mathcal{G}(s,t)$$

with initial condition $\mathcal{G}(s,0) = s^x$. The expected behaviour is given by $\frac{\partial \mathcal{G}}{\partial s}(1,t)$ and the probability of absorption is given by $\mathcal{G}(0,t)$. Taking the limit as $t \to \infty$ illustrates the fact that extinction is certain precisely when $\gamma \geq \beta$.

Corollary 3.2.2 (Anderson [1991]). *For the Linear BDP with β, $\gamma > 0$ we have*

$$\mathbb{P}_x[T < \infty] = \begin{cases} 1 & \text{if } \beta \leq \gamma \\ (\gamma/\beta)^x & \text{otherwise.} \end{cases}$$

As in the case of the SIS and SIR epidemic models defined in Section 2.2, there are two main cases to consider: when $\beta > \gamma$ and when $\beta < \gamma$. We will refer to these two regimes as *supercritical* and *subcritical* respectively; the critical case is when $\beta = \gamma$. In the supercritical case, we have seen in Corollary 3.2.2 that extinction is not certain and so by Theorem 2.1.16, a QSD cannot exist. In the subcritical and critical cases, we need to calculate $\mathcal{A}$ in (3.1.1) in order to apply Theorem 3.1.1 to the Linear BDP. By substituting in the linear birth and death rates, we obtain, as in Collet et al. [2001],

$$\pi_n = \left(\frac{\beta}{\gamma}\right)^{n-1} \frac{1}{n}$$

and hence the series (3.1.1) becomes

$$\mathcal{A} = \sum_{n=1}^{\infty} \frac{1}{\beta_n \pi_n} \sum_{i=n+1}^{\infty} \pi_i = \sum_{n=1}^{\infty} \frac{\gamma^{n-1}}{\beta^n} \sum_{i=n+1}^{\infty} \frac{\gamma^{i-1}}{i\beta^{i-1}}$$

When $\beta \leq \gamma$, this is bounded below by

$$\sum_{n=1}^{\infty} \frac{\gamma^{n-1}}{\beta^n} \sum_{i=n+1}^{\infty} \frac{1}{i} \leq \mathcal{A}$$

and so $\mathcal{A}$ diverges as each of the summands diverges. The final component we require in order to apply Theorem 3.1.1 is the evaluation of the decay parameter α for the Linear BDP. Through applications of spectral theory and orthogonal polynomials, van Doorn [1991] Example 6.1 finds that $\alpha = \gamma - \beta$ when $\beta < \gamma$. This allows us to apply Theorem 3.1.1 to arrive at the following result, as summarised in van Doorn and Pollett [2013]:

Theorem 3.2.3 (Theorem 16, van Doorn and Pollett [2013]). *Let $X = (X(t))_{t \geq 0}$ be a linear birth death process with birth rate $\beta > 0$ and death rate $\gamma > 0$. When $\beta \geq \gamma$ there is no QSD for X. When $\beta < \gamma$ there exists an uncountable collection of QSDs for X: $\{\mathbf{u}(x) : 0 < x \leq \alpha\}$.*

3.2.1 LCDs for finite-support initial conditions

If we restrict to the case where the initial distribution has finite support, then van Doorn [1991] arrives at the following result which we state for the LCD starting from a single infective, but which holds for any initial distribution with finite support.

Proposition 3.2.4 (Example 6.1, van Doorn [1991])**.** *Let $(X(t))_{t\geq 0}$ be a Linear BDP with parameters $\beta > \gamma > 0$, and initial condition $X(0) = 1$. Then the 1-LCD is given by*

$$u_i = \left(\frac{\beta}{\gamma}\right)^{i-1}\left(1 - \frac{\beta}{\gamma}\right) \qquad i \geq 1 \tag{3.2.4}$$

and so is a geometric distribution with parameter $R_0 = \beta/\gamma$.

This follows from the α-invariance property ($-\alpha\mathbf{u}^T = \mathbf{u}^T Q$), where $\alpha = \gamma - \beta$ by Theorem 3.2.3, and then by performing induction on the expression for u_i.

3.2.2 Characterising x-invariant QSDs and LCDs

Section 3.2.1 fully characterises the $\mathbf{v}$-LCD starting from distributions $\mathbf{v}$ with finite support, but we would like to extend this to the rest of the collection of QSDs, and to which LCDs they correspond. I adapt the iterative expression of $\mathbf{u}(\alpha)$ to obtain iterative expressions for $\mathbf{u}(x)$ with $0 < x < \alpha$.

Writing out the x-invariance system of equations we get that $\mathbf{u}(x) = (u_i)_{i\geq 1}$ satisfies

$$\begin{aligned} u_1 &= \frac{x}{\gamma} \\ u_2 &= \frac{((\beta+\gamma) - x)u_1}{2\gamma} \\ u_{i+1} &= \frac{(i(\beta+\gamma) - x)u_i - (i-1)\beta u_{i-1}}{\gamma(i+1)} \quad i \geq 2 \end{aligned} \tag{3.2.5}$$

The value for u_1 follows from Theorem 2.1.13.

Although these cannot be solved analytically, one can easily implement these numerically, and hence in Figure 3.2.1 we see examples of $\mathbf{u}(x)$ for some values of $0 < x < \alpha$. This diagram illustrates the fact that the lower x is, the heavier the tail of the distribution.

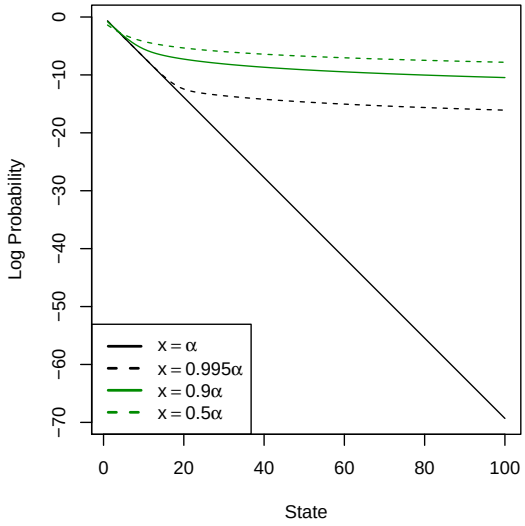


Figure 3.2.1: Comparison of QSDs $\mathbf{u}(x)$ for Linear BDP with $\beta = 0.5$, $\gamma = 1$ as x varies. Plotted on logarithmic scale for clarity.

Proposition 3.2.5. *Let $0 < x < \alpha$. Then the probability generating function for $\mathbf{u}(x)$ is given by*

$$\mathcal{G}(s) = 1 - \left(\frac{\gamma(1-s)}{\gamma - \beta s}\right)^{\frac{x}{\gamma-\beta}} \tag{3.2.6}$$

Consequently, $\mathbf{u}$ is a proper distribution and the mean of $\mathbf{u}(x)$, $\sum_{i=1}^{\infty} iu_i = \infty$.

Proof. In order to show $\mathbf{u}(x)$ is a proper distribution, we derive the generating function. Let $\mathcal{G}(s) = \sum_{k=1}^{\infty} u_k s^k$. To obtain an expression for $\mathcal{G}(s)$ we consider the system of equations (3.2.5), and multiply through by s_k for the equation with left-hand side u_k. Summing over k and rearranging gives the differential equation: $xs(1 - \mathcal{G}(s)) = \mathcal{G}'(s)(\gamma s - (\beta + \gamma)s^2 + \beta s^3)$. With the initial condition $\mathcal{G}(0) = 0$, we solve this differential equation to obtain the expression in (3.2.6). We see that $\mathcal{G}(1) = 1$ so the distribution is proper, and

$$\mathcal{G}'(s) = \frac{-x}{\gamma - \beta}\left(\frac{\gamma(1-s)}{\gamma - \beta s}\right)^{-1+x/(\gamma-\beta)}\left(\frac{\gamma(\beta-\gamma)}{(\beta s - \gamma)^2}\right)$$

For this expression $\mathcal{G}'(1^-) = \lim_{s\uparrow 1} \mathcal{G}'(s) = \infty$, so the mean is infinite for any $x < \alpha$. □

3.3 Pure Death Processes

One special case of the birth-death process is when we set all the birth rates β_i equal to zero. For the pure death process we denote the sequence of strictly positive death rates by $(\gamma_i)_{i\geq 1} > 0$ as before. This means that the extinction time $T < \infty$, since from any initial state, the lifetime is the finite sum of a sequence of exponential variables $T_i < \infty$. This simplification will allow us to obtain more specific results regarding the characterisation of the QSDs and the LCDs for the death process. Once again, the simplest case is the linear death process, where $\gamma_i = i\gamma$ for all i, modelling a population with independent $\text{Exp}(\gamma)$ lifetimes.

One application of such a death process could be in the modelling of a competing collection of pathogens within an environment. Under the right circumstances, one would expect to see many of the competing pathogens to die out over time, leaving the strongest (or luckiest). To put this into a more probabilistic framework, one could consider a multi-type process for which, once a type has gone extinct it cannot return. Then under some assumptions on the linearity of the model, one might be able to couple such a process with a death process to more simply model the number of different pathogens in the environment.

In this section, we will refine the results from Section 3.1, and fully characterise the QSDs in the case of the linear death process, in both the finite and countable settings. In addition, we will fully characterise the finite-support LCDs for a general death process, and discuss the infinite-state space case.

3.3.1 The Linear Death Process

We first look at the special case of linear pure death processes where deaths occur with rate $\gamma > 0$. If we assume the initial condition has finite support contained in $\{1, 2, \ldots, L\}$ for some $L \geq 1$, then we can assume without loss of generality that the process evolves on a finite state space $S \cup \{0\}$ where $S = \{1, \ldots, L\}$. As a consequence, we can make use of the results in Section 2.1.1 to characterise the process. Firstly, we note that each state is in its own communicating class $S_l = \{l\}$; this satisfies the ordering convention in Section 2.1.1.

Proposition 3.3.1. *The linear death process $X = (X(t))_{t\geq 0}$ with $X(0) = L$ on state space $\bar{S} = \{1, \ldots, L\} \cup \{0\}$ with death rate $\gamma > 0$ has a unique QSD which*

corresponds to the $\mathbf{v}$*-LCD for any proper probability distribution* $\mathbf{v}$*. This QSD gives full weight to state 1.*

Proof. To use Theorem 2.1.11, we compute the decay parameter α_i on leaving class $S_i = \{i\}$. To do so, we need only look at the Q-matrix restricted to that class, and hence we need only consider $q_{ii} = -\gamma i$. This means that if α_i is the decay parameter on leaving class S_i, then $\alpha_i = \gamma i$. We can then obtain the decay parameter α for the entire process by finding the minimum: $\alpha = \min\{\alpha_i : i \in S\} = \min\{i\gamma : i \in S\} = \gamma$.

Using the notation in Section 2.1.1, we see that the decay parameter $\alpha = \gamma$ corresponds uniquely to class $S_1 = \{1\}$. Hence $S_{\min} = S_1$ which is always minimal for α. Using Theorem 2.1.11, we get that there is a unique QSD from which $\{1\}$ is accessible. This QSD only gives weight to states accessible from state 1, and since only $\{0\}$ is accessible from $\{1\}$ we must have that $u_i = 0$ for all $i \geq 2$. □

We extend this notion to include a much wider range of initial distributions, ones with infinite support. In this case, we can consider the death process as evolving on a countable state space $S = \mathbb{N}$. When we extend to the countable case, we still retain the finite support QSD $\mathbf{u}(\alpha)$ given above, but we can also consider the uncountable collection of QSDs which correspond to $\mathbf{v}$-LCDs where $\mathbf{v}$ has infinite support. Since we have the equivalence between x-invariance and quasi-stationarity in Theorem 2.1.13 , we can obtain the following result.

Theorem 3.3.2. *Let* $(X(t))_{t\geq 0}$ *be a linear death process with death rate* $\gamma > 0$*. Then for each* $x \in (0, \gamma]$*, there exists a QSD* $\mathbf{u}(x)$ *given by*

$$\begin{aligned} u_1 &= x\gamma^{-1} \\ u_n &= \frac{x}{\gamma n!}\prod_{i=2}^{n}((i-1) - x/\gamma) \qquad n \geq 2 \end{aligned} \tag{3.3.1}$$

and this distribution is proper.

This follows straight from the x-invariance equations (3.2.5) and the generating function in Proposition 3.2.5, setting $\beta = 0$. Note that this matches the result in Theorem 2.1.13, which states that in the irreducible case, we must also have $u_1 = x\gamma^{-1}$.

As we can see in Figure 3.3.1, these behave similarly to the linear birth-death process where the QSDs corresponding to $x < \alpha$ have heavier weighting away from zero.

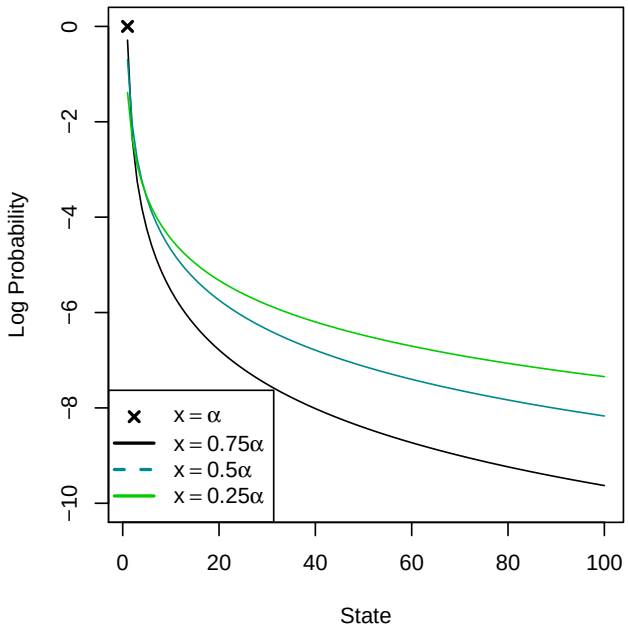


Figure 3.3.1: Comparison of QSDs $\mathbf{u}(x)$ for Linear Pure Death Process as x varies for $x \in (0, \alpha]$. Plotted on a logarithmic scale for clarity.

Indeed, we prove that, even in the pure death case, the QSDs for $x < \alpha$ have infinite mean.

Proposition 3.3.3. *For the linear pure death process with $\gamma > 0$, every x-invariant QSD $\mathbf{u}(x)$ with $x < \gamma$ has infinite mean.*

Proof. As in Theorem 3.3.2 we know the generating function $\mathcal{G}(s) = -\frac{x}{\gamma}(1-s)^{x/\gamma}+1$. We then observe for $x/\gamma < 1$ that:

$$\lim_{s\to 1} \mathcal{G}'(s) = \lim_{s\to 1} \frac{x^2}{\gamma^2}(1-s)^{x/\gamma-1} = \infty. \qquad \square$$

3.3.2 General Death Processes

We now extend our process to the case with general death rates. If we return to the finite case, where the process evolves on $\{1, \ldots, L\}$ we obtain existence of a QSD by Theorem 2.1.10. Indeed, one can easily see that the probability measure giving full weight to state 1 is still a QSD in the general case; think of it as the 1-LCD. On the other hand, the claim of uniqueness cannot hold in general. Indeed, consider the following example.

Example 3.3.4. Consider the death process $(X(t))_{t\geq 0}$ on $S = \{1, \ldots, 5\}$ with the death rates $\gamma_1 = 3$, $\gamma_2 = 2$, $\gamma_3 = 3$, $\gamma_4 = 1$, $\gamma_5 = 3$. If we consider the initial condition $X(0) = 5$, then by solving the x-invariance equation, we obtain the QSD given by $\mathbf{u}(x) = (1/3, 1/3, 1/9, 2/9, 0)$. For $X(0) = 4$ we obtain the same result. However, one can also show that $(1, 0, 0, 0, 0)$ is still a QSD for this process, and so uniqueness doesn't hold.

In order to better characterise the QSDs of the process, we instead consider the LCDs starting from different initial conditions. For a fixed starting position j, we know the j-LCD exists and is unique, but more generally, we would like to know for which pairs (i, j) the i-LCD and j-LCD coincide. We take the known results regarding the existence of the minimal QSDs on a reducible state space in Theorem 2.1.11, and from van Doorn and Pollett [2008] and extend them for the death process to fully characterise the set of j-LCDs.

Theorem 3.3.5. *Let $(X(t))_{t\geq 0}$ be a general death process on $S \cup \{0\} = \{0, \ldots, L\}$ with rates $\{\gamma_1, \ldots, \gamma_L\}$ ($\gamma_i > 0$ for all $i = 1, \ldots, L$). For $j = 1, \ldots, L$ define $\gamma^*(j) = \min\{\gamma_i : i \leq j\}$, and let $i^*(j) = \min\{i \leq j : \gamma_i = \gamma^*\}$. Then the j-LCD gives weight precisely to the states $\{1, \ldots, i^*(j)\}$.*

Proof. For each $j = 1, \ldots, L$, we can consider the process starting from j to be a process which evolves on $\{0, \ldots, j\}$ only without loss of generality. Then we have the decay parameter for the process must be $\alpha = \gamma^*(j)$. If $\gamma_i \neq \gamma^*(j)$ for all $i \neq i^*(j)$, then α has geometric multiplicity one and therefore, by Theorem 2.1.11, there exists a QSD which is $\gamma^*(j)$-invariant for Q and gives weight to all the states $\{1, \ldots, i^*(j)\}$. This is because state $i^*(j)$ is minimal for $\alpha = \gamma^*(j)$. Furthermore, we must have that this is the j-LCD by the same theorem. If $\gamma_k = \gamma^*(j)$ for some $k \leq j$ with $k \neq i^*(j)$, then straight away it must be the case that $k > i^*(j)$ by definition of $i^*(j)$. Moreover, we can see from the x-invariance equations, that under $\gamma^*(j)$-invariance we must have

$$-\gamma^*(j)u_{i^*(j)} = -\gamma^*(j)u_{i^*(j)} + \gamma_{i^*(j)+1}u_{i^*(j)+1}$$

which forces $u_{i^*(j)+1} = 0$, and henceforth $u_i = 0$ for $i > i^*(j)$, and in particular u_k. Therefore, despite the multiplicity, we still have uniqueness of the $\gamma^*(j)$-invariant QSD. □

Intuitively, we think of the smallest values of γ_i as "bottlenecks", where the process

gets hung up on during its progression to zero. What this result tells us is that from any initial starting location j, the QSD gives weight up to the tightest bottleneck below the starting location. In the linear case (Section 3.3.1), the only bottleneck is in state 1, so all LCDs are the same. In Example 3.3.4 above, we have bottlenecks at states 2 and 4, so the 3-LCD gives weight to states 1 and 2, and the 5-LCD gives weight to the first four states.

Note that in the above theorem, this can be extended immediately to include other initial distributions with finite support. For initial distribution $\mathbf{v}$ with finite support, if $j_{\mathbf{v}} = \max \operatorname{supp}(\mathbf{v})$ then the $\mathbf{v}$-LCD is exactly the $j_{\mathbf{v}}$-LCD.

Note that we can use Theorem 3.3.5 to write down the full list of QSDs for the finite death process.

Corollary 3.3.6. *For the general death process on L states with death rates $(\gamma_i)_{i=1}^{L}$, let $A = \{1 \leq i \leq L : \gamma_i < \gamma_j \text{ for all } j < i\}$. Then the full set of QSDs for the death process is precisely the set of i-LCDs for $i \in A$. Each i-LCD is $\gamma^*(i)$-invariant for Q.*

The above theorem and corollary fully characterise the QSDs and LCDs for the general death process on a finite state space. We note that since these exist on the finite state space there are only finitely many such distributions.

Countable State Space

If we now consider the countable case, then we have death rates $(\gamma_i)_{i=1}^{\infty}$. One can apply Theorem 3.3.5 to obtain the finite support LCDs.

Corollary 3.3.7. *Let $(X(t))_{t\geq 0}$ be a general death process on $\mathbb{N} \cup \{0\}$ with death rates $\gamma_i \geq c > 0$ for some $c > 0$. Then for each $j \geq 1$ there exists a j-LCD which gives weight precisely to states $\{1, \ldots, i^*(j)\}$ and which is $\gamma^*(j)$-invariant. This j-LCD is also equal to the $\mathbf{v}$-LCD for any $\mathbf{v}$ with $\max(\operatorname{supp}(\mathbf{v})) = j$.*

This holds since for any $j > 0$ we can consider the process to be evolving just on $\{1, \ldots, j\}$ and this returns us to the finite case. We can however extend this result to take advantage of the countable setting. The following result is fairly general, but can be applied in the special case of the linear death process and generate all of the QSDs.

Theorem 3.3.8. *Let $(X(t))_{t\geq 0}$ be a general death process on $\mathbb{N}\cup\{0\}$ with death rates $c<\gamma_i<C$ strictly bounded away from zero and infinity by c, C independent of i, and attaining the infimum, $\gamma^*=\inf_j \gamma_j$, so $\gamma_k=\gamma^*$ for some k. Then for each $x\in(0,\gamma^*)$ there exists an x-invariant QSD which is a proper probability distribution giving mass to all states in $\mathbb{N}$.*

Proof. Note that in this case, since we are considering the countable state space, we need only consider the smallest bottleneck, which we force to exist by construction. In this case, we can iteratively solve the x-invariance equations, which all give positive solutions since they are of the form

$$u_{j+1}=\frac{\gamma_j-x}{\gamma_{j+1}}u_j$$

where $\gamma_j-x>0$ for all $x<\gamma^*$. This can then be normalised since

$$u_{j+1}=\frac{\gamma_1-x}{\gamma_{j+1}}\prod_{i=2}^{j}\frac{\gamma_i-x}{\gamma_i}u_1<\frac{\gamma_1-x}{\gamma^*}\left(1-\frac{x}{C}\right)^{j-1}u_1$$

This follows since $0<\gamma_j-x<\gamma_j$. This boundedness shows boundedness above of $\sum_i u_i$ by the geometric series, and hence we can renormalise the sum to obtain the expression for the QSD. □

3.4 SIS model as a birth-death process

The final section of this chapter looks at the SIS model as a third form of birth-death process, one on a finite state space. Indeed, one can consider the SIS epidemic model to be a birth-death process with birth rate $\beta_i = \beta i(N-i)/N$ for $i < N$, and death rates $\gamma_i = i\gamma$. As long as the initial distribution has support contained in $\{1, \ldots, N\}$, this is well defined. It gives us a finite state space with a single transient communicating class, and hence we have uniqueness of the QSD in the SIS model for all values of $\beta, \gamma > 0$. However, when one attempts to compute the QSD for the SIS model it turns out that there is no closed-form expression for the QSD. Furthermore, we cannot even give a closed expression for the decay parameter for this model in general.

As such, one turns to approximations. Much work by Nåsell [1999] was done on approximations using normal distributions and geometric distributions, depending on whether R_0 is much bigger or smaller than 1, taking into account the population size. This is also linked to work by Kryscio and Lefèvre [1989] which considers two approximating processes: one with a reflecting boundary at 0, and one with an "immortal" infective, who remains infective for all time. We will consider two different approaches which generalise these, and observe the stationary distributions for the approximating models and how they compare to the true QSD, which for small populations can be computed numerically through the x-invariance equations.

3.4.1 Externally Restarted SIS Model

Firstly, we include an external infective source which restarts the epidemic after it goes extinct (on the event $I(t) = 0$). At this point we wait for an exponentially-distributed time with rate $\rho > 0$, at which point we infect a single individual uniformly at random. Other than this, the epidemic progresses as before. More specifically,

$$\mathbb{P}[I(t + \Delta t) = 1 | I(t) = 0] = \rho \Delta t + o(\Delta t)$$

where $S(t) = N - I(t)$ for all $t \geq 0$. We will refer to this model as (SIS-A). Given this construction, we can simply compute the stationary distribution for the process.

Theorem 3.4.1. *For (SIS-A) as described above, the stationary distribution is given by:*

$$\begin{aligned}
\pi_i &= \binom{N}{i}(i-1)!\frac{\beta^{i-1}}{\gamma^{i-1}N^i}\pi_1 && i \geq 2 \\
\pi_0 &= \frac{\gamma}{\rho}\pi_1 \\
\pi_1 &= \left(\frac{\gamma}{\rho} + \sum_{i=1}^{N}\binom{N}{i}(i-1)!\frac{\beta^{i-1}}{\gamma^{i-1}N^i}\right)^{-1}
\end{aligned}$$

This follows directly from solving the system of equations $\pi^T Q = \mathbf{0}$ and normalising the solutions, which we know to be a finite sum. By using this description, we can clearly see that the relative weight of states $\{1, \ldots, N\}$ do not depend on ρ, only π_0 and π_1 depend on ρ directly.

To get an idea of the endemic level of infection in the population in the (SIS-A) model we look at the modes of the stationary distribution.

Theorem 3.4.2. *For the stationary distribution of (SIS-A), if $\rho < \gamma$ there is a mode at 0. In the case where $R_0 > 1$ and $N > R_0/(R_0-1)^2$, there is a second mode $\lceil m \rceil$ where m is given by*

$$\frac{m}{N} = \frac{\sqrt{R_0^2 - 4R_0 - 4R_0/N + 1} + R_0 - 1}{2R_0} \to 1 - \frac{1}{R_0} \text{ as } N \to \infty$$

Proof. We consider the ratio π_{k+1}/π_k and look for the point $k \geq 1$ where the ratio equals one; this is similar to looking at the gradient of the distribution in a continuous setting. Computing this we get that (for a population of size N):

$$\frac{\pi_{k+1}}{\pi_k} = \frac{\binom{N}{k+1}k!R_0^k N^{-(k+1)}}{\binom{N}{k}(k-1)!R_0^{k-1}N^{-k}} = \frac{k(N-k)R_0}{N(k+1)} \qquad k \geq 1.$$

We note that $\pi_0 > \pi_1$ in all cases and obtain the mode at 0. For $k > 0$, solving the inequality $\pi_{k+1}/\pi_k > 1$ is equivalent to $R_0k^2 + (N - R_0N)k + N < 0$. We look at the positive roots of this equation to get

$$m^{\pm} = \frac{\pm\sqrt{R_0^2N^2 - 2R_0N^2 - 4R_0N + N^2} + R_0N - N}{2R_0}$$

and, after simplifying the expression, the result follows. □

We note from this result that this does not depend on ρ, the restarting parameter. This suggests that one can choose ρ to best aid simulation of the method: larger ρ cause a faster restart of the process, meaning less "dead" time for any simulations you might wish to run for it. We also observe that the same threshold number holds for (SIS-A) as for the original SIS model: we only obtain a non-zero mode for $R_0 > 1$, and in this case we see the two possibilities of a large or a small outbreak relative to the size of the population. This has recently been considered in Section 3 of Da et al. [2016]

3.4.2 Constant External Infective Pressure

We now consider a second modification to the SIS model where we include a constant external infective source. This can be thought of as an averaged effect of individuals being infected from sources outside the population: either other people or different sources. In this case, in addition to the internal infections with rate $\beta S(t)I(t)/N$, each susceptible individual is independently infected by an external source according to a Poisson process with rate $\rho > 0$. More specifically,

$$\mathbb{P}[I(t+\Delta t) = i+1 | I(t) = i] = \rho(N-i)\Delta t + \beta i \frac{(N-i)}{N}\Delta t + o(\Delta t) \qquad i > 0$$

We shall refer to this model with constant external infective pressure as (SIS-B).

We can also obtain the exact expression for the stationary distribution for this process.

Theorem 3.4.3. *For model (SIS-B) as described above, we have the stationary distribution is given by:*

$$\pi_i = \binom{N}{i} \sum_{j=0}^{i-1} \left(\frac{\rho^{i-j}\beta^j}{\gamma^i N^j} \mathcal{B}(i-1, j) \right) \pi_0 \qquad i \geq 1$$

$$\pi_0 = \left(1 + \sum_{i=1}^{N} \binom{N}{i} \sum_{j=0}^{i-1} \frac{\rho^{i-j}\beta^j}{\gamma^i N^j} \mathcal{B}(i-1, j) \right)^{-1}$$

where

$$\mathcal{B}(k, l) = \sum_{a \in \Omega(k,l)} \prod_{i=1}^{l} a_i$$

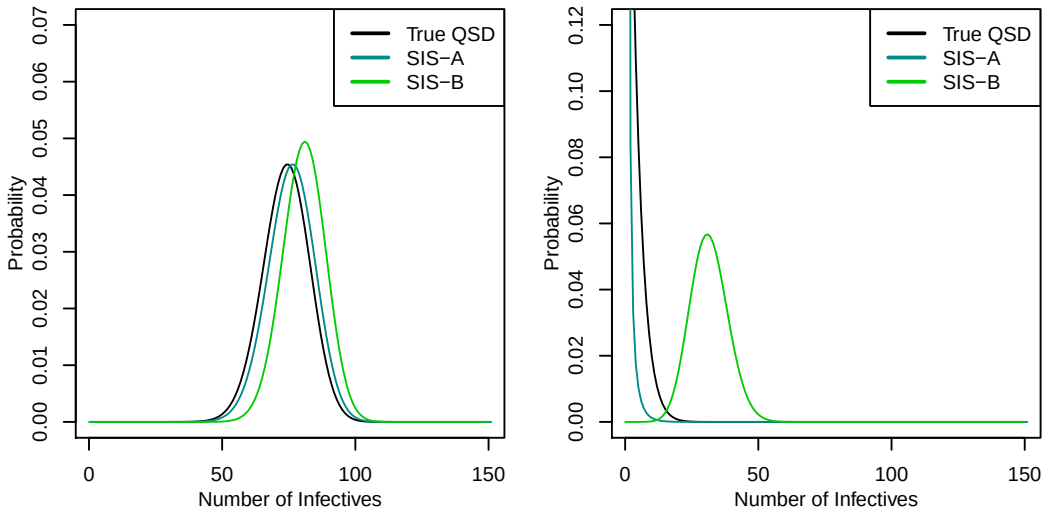
and $\Omega(k, l)$ is the set of all l-tuples $a = (a_1, \ldots, a_l)$ one can draw from $\{1, \ldots, k\}$ without replacement. We also define $\mathcal{B}(k, 0) = 0, \mathcal{B}(0, l) = 1$.

The proof, which can be found in Appendix A.3, again solves $\pi^T Q = \mathbf{0}$, and makes use of the recurrence relation $\mathcal{B}(i+1,j) = \mathcal{B}(i,j) + (i+1)\mathcal{B}(i,j-1)$ for $0 < j \leq i$. Note that $\mathcal{B}(j,k)$ can be potentially computationally intensive for large populations, either in storing precomputed values (memory issues) or in inline computations (time intensive). However, one may be able to use the recurrence relations in order to speed up such computation.

3.4.3 Comparison of Approximating Processes

We see in Figure 3.4.1 that the two processes perform differently in different cases. In the subcritical case, we see clearly that (SIS-A) gives a much closer approximation, since (SIS-B) still gives a stationary distribution with a non-zero mode. Even in the supercritical case, we see that (SIS-A) performs better, but (SIS-B) still gives a close answer.

Our examples above are variations of two existing approximations in Nåsell [1999] in obtaining the QSD for the standard SIS epidemic model. The first has reflection away from state 0 rather than absorption, akin to the limiting behaviour of (SIS-A) as $\rho \to \infty$. This second uses a single permanently infected individual (equivalent to our (SIS-B) with $\rho = \beta$). We see in Figure 3.4.2 that, in the subcritical case, (SIS-B) improves as ρ goes to zero, since the external infection pressure has less effect on the process, but still keeps it recurrent. It would be of interest to consider equivalent approximations for the SIRS model, and look for exact solutions and convergence results.



(a) $\beta = 2$, $\gamma = 1$, $N = 150$, $\rho = 0.1$ (b) $\beta = 0.8$, $\gamma = 1$, $N = 150$, $\rho = 0.1$

Figure 3.4.1: Comparisons of approximating stationary distributions with the true QSD

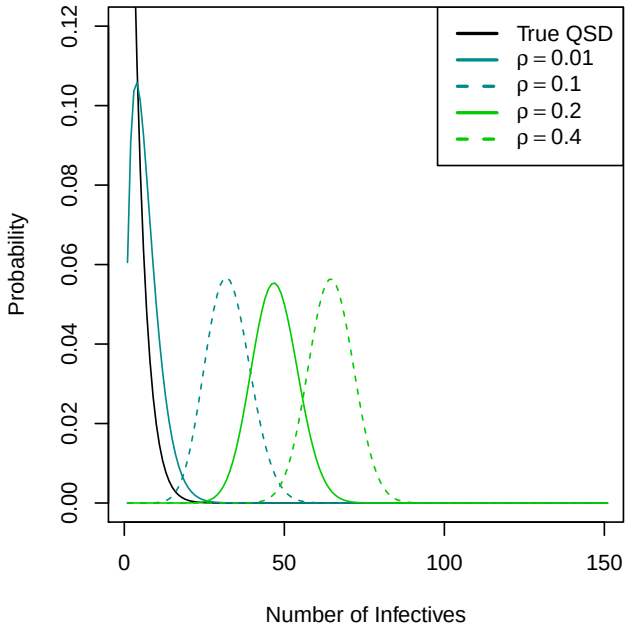


Figure 3.4.2: Observation of accuracy of approximations as ρ changes for (SIS-B) in the subcritical case $\beta = 0.8$, $\gamma = 1$, $N = 150$.

Chapter 4

Characterisation of the Transient Immunity Process

In this chapter, we take our first step to extend the birth-death process. As discussed in Chapter 3, the Linear BDP can be compared to the SIS epidemic model evolving in an infinite susceptible population, serving as a good method of approximation via a linear model in large enough populations. Given the evolving nature of pathogens such as influenza, one natural question to ask is whether it is appropriate to assume that an individual can be reinfected immediately following recovery, as in the SIS model. One might wish consider an alternative where, following an infectious period, an individual develops some level of immunity to the pathogen. If this immunity was permanent, it would lead us to consider the SIR model. However, through means such as antigenic drift the developed immunity may become less and less effective or relevant against strains which emerge further in the future. To this end, we initially consider a simpler mechanism where an individual has total immunity from all infection for some period of time, and then becomes totally susceptible to infection again. This leads us to consider the SIRS epidemic model defined in Section 2.2.1, where, following recovery, an individual is globally immune to reinfection for an exponentially-distributed period of time. As in Section 3.4, we want to consider a linear process which approximates the SIRS model for sufficiently large populations. To this end, we extend the Linear BDP to include a "recovery" period; this will be called the *Transient Immunity Process*.

This chapter will start with an outline of the Transient Immunity Process and

some basic properties of the process, such as the expected behaviour. Following this, we will investigate the different notions of quasi-stationarity related to this model: conditioning on still having a positive number of infectives, or conditioning on still having either immune or infective individuals in the population. Following this, we will discuss existence and uniqueness results for each of these versions of quasi-stationarity. In addition we investigate how the act of conditioning affects the asymptotic expected behaviour of the process.

4.1 Construction and Behaviour

We define the *Transient Immunity Process* as follows. The Transient Immunity process $(\mathbf{Y}(t))_{t\geq 0} = (I(t), R(t))_{t\geq 0}$ is a bivariate Markov process taking values in $\mathbb{N}_0 \times \mathbb{N}_0$ characterised by the following events for given *infection rate* $\beta > 0$, *recovery rate* $\gamma > 0$ and *loss of immunity rate* $\delta > 0$:

Infection Event:

$$\mathbb{P}[\mathbf{Y}(t+\Delta t) = (i+1, r)|\mathbf{Y}(t) = (i, r)] = \beta i \Delta t + o(\Delta t) \qquad i \geq 1, r \geq 0$$

Recovery Event:

$$\mathbb{P}[\mathbf{Y}(t+\Delta t) = (i-1, r+1)|\mathbf{Y}(t) = (i, r)] = \gamma i \Delta t + o(\Delta t) \qquad i \geq 1, r \geq 0$$

Loss of Immunity Event:

$$\mathbb{P}[\mathbf{Y}(t+\Delta t) = (i, r-1)|\mathbf{Y}(t) = (i, r)] = \delta r \Delta t + o(\Delta t) \qquad i \geq 0, r \geq 1$$

No event in $(t, t+\Delta t)$:

$$\mathbb{P}[\mathbf{Y}(t+\Delta t) = (i, r)|\mathbf{Y}(t) = (i, r)] = (1 - (\beta i + \gamma i + \delta r)\Delta t) + o(\Delta t) \qquad i \geq 0, r \geq 0$$

We note that $\mathbf{Y}(t)$ is a reducible process which evolves over the state space $\bar{S} = \mathbb{N}_0 \times \mathbb{N}_0$. We can decompose $\bar{S}$ into the following partition:

$$\bar{S} = S \cup \{(0,0)\} = S^* \cup \{(0, r) : r > 0\} \cup \{(0,0)\}$$

where we use the notation

$$\begin{aligned} S &:= \mathbb{N}_0 \times \mathbb{N}_0 \setminus \{(0,0)\} \\ S^* &:= \{(i,r) : i > 0, r \geq 0\} \end{aligned} \tag{4.1.1}$$

for the whole transient space and the countably infinite communicating class, and $S_r = \{(0,r)\}$ for each single-state communicating class.

4.1.1 Expectation and Variance

Firstly, we will look at the expected behaviour of $I(t)$ and $R(t)$. The expected behaviour of $I(t)$ has already been shown in Theorem 3.2.1 since, by observing the construction above, one can see that the I-component doesn't depend on $R(t)$ and has the same behaviour as the Linear BDP. Indeed, if we define

$$\begin{aligned} T_i^I &:= \inf\{t \geq 0 : I(t) = 0 | I(0) = i\} \\ T_{(i,r)}^{(TI)} &:= \inf\{t \geq 0 : (I(t), R(t)) = (0,r) \text{ for some } r \geq 0 | (I(0), R(0)) = (i,r)\} \end{aligned}$$

then we can also conclude that we have equality in distribution

$$T_i^I = T_{(i,r)}^{(TI)} \qquad \text{for all } i \geq 1, r \geq 0. \tag{4.1.2}$$

Theorem 4.1.1. *For the Transient Immunity Process $\mathbf{Y}(t) = (I(t), R(t))_{t \geq 0}$ with initial conditions $I(0) = i_0, R(0) = r_0$ and parameters $\beta, \gamma, \delta > 0$ we have that, for $t \geq 0$, the expected behaviour is given by:*

$$\mathbb{E}[I(t)] = \begin{cases} i_0 e^{(\beta-\gamma)t} & \beta \neq \gamma \\ i_0 & \beta = \gamma \end{cases}$$

$$\mathbb{E}[R(t)] = \begin{cases} \frac{i_0 \gamma}{(\beta-\gamma)+\delta} \left(e^{(\beta-\gamma)t} - e^{-\delta t} \right) + r_0 e^{-\delta t} & \gamma - \beta \neq 0, \gamma - \beta \neq \delta \\ \frac{\gamma i_0}{\delta}(1 - e^{-\delta t}) + r_0 e^{-\delta t} & \gamma - \beta \neq 0, \gamma - \beta = \delta, \\ (\gamma i_0 t + r_0)\, e^{-\delta t} & \gamma - \beta = 0, \gamma - \beta \neq \delta \end{cases}$$

This theorem follows from standard differential equation solutions.

In these above cases, we see three types of behaviour exhibited depending on the value of $(\beta - \gamma)$, or equivalently the reproductive number $R_0 = \beta/\gamma$.

- If $(\beta - \gamma) < 0$, then we get $\mathbb{E}[R(t)] \to 0$ like $O(e^{(\beta-\gamma)t} \wedge e^{-\delta t})$.
- If $(\beta - \gamma) > 0$, then we get $\mathbb{E}[R(t)] \to \infty$, regardless of the size of δ.
- If $(\beta - \gamma) = 0$, then we get convergence to zero like $\mathbb{E}[R(t)] = O(te^{-\delta t})$.

In addition to the expected behaviour we mention the variance of the components of the process, which we will compare to the variance of the components of the process under the different forms of conditioning.

Theorem 4.1.2. *The variance of the number of infectives for the Transient Immunity Process, Var$(I(t))$, with initial conditions $\mathbf{Y}(0) = (1, 0)$ is given by*

$$Var(I(t)) = \begin{cases} \frac{\beta+\gamma}{\beta-\gamma} i_0 \left(e^{2(\beta-\gamma)t} - e^{(\beta-\gamma)t} \right) & \beta \neq \gamma \\ (\beta + \gamma) i_0 t & \beta = \gamma \end{cases} \tag{4.1.3}$$

Proof. In order to prove the theorem, we will determine the differential equation for the second moment $\mathbb{E}[I(t)^2]$, solve it and substitute into the expression for $\mathrm{Var}(I(t))$. Given we only need to consider infection and recovery events, we note that

$$I(t+h) = I(t) \pm 1 \Rightarrow I(t+h)^2 - I(t)^2 = (I(t)^2 \pm 2I(t) + 1) - I(t)^2 = 1 \pm 2I(t) \tag{4.1.4}$$

Conditioning on the next event at time t and on $I(t)$ we obtain

$$\mathbb{E}[I(t+h)^2 - I(t)^2] = \mathbb{E}\left[\mathbb{E}[I(t+h)^2 - I(t)^2 | I(t)]\right]$$

Taking the inner expectation and using (4.1.4) gives

$$\begin{aligned} \mathbb{E}[I(t+h)^2 - I(t)^2] &= \mathbb{E}\left[(1 + 2I(t))\beta h I(t) + (1 - 2I(t))\gamma h I(t)\right] + o(h) \\ &= 2(\beta - \gamma)h\mathbb{E}[I(t)^2] + (\beta + \gamma)h\mathbb{E}[I(t)] + o(h) \\ &= 2(\beta - \gamma)h\mathbb{E}[I(t)^2] + (\beta + \gamma)h i_0 e^{(\beta-\gamma)t} + o(h) \end{aligned}$$

using $\mathbb{E}[I(t)] = i_0 e^{(\beta-\gamma)t}$. Dividing through by h and taking the limit as h goes to 0 gives the differential equation

$$\frac{d}{dt}\mathbb{E}[I(t)^2] = 2(\beta - \gamma)\mathbb{E}[I(t)^2] + (\beta + \gamma) i_0 e^{(\beta-\gamma)t} \tag{4.1.5}$$

We then solve (4.1.5) by use of an integrating factor and the initial conditions

$\mathbb{E}[I(0)^2] = i_0^2$ to give

$$\mathbb{E}[I(t)^2] = \frac{\beta+\gamma}{\beta-\gamma} i_0 \left(e^{(\beta-\gamma)t} - e^{2(\beta-\gamma)t}\right) + i_0^2 e^{2(\beta-\gamma)t} \tag{4.1.6}$$

Putting (4.1.6) into the expression $\mathrm{Var}(I(t)) = \mathbb{E}[I(t)^2] - \mathbb{E}[I(t)]^2$ gives

$$\mathrm{Var}(I(t)) = \frac{\beta+\gamma}{\beta-\gamma} i_0 \left(e^{(\beta-\gamma)t} - e^{2(\beta-\gamma)t}\right)$$

proving the first case. The second case follows using direct integration of both sides of (4.1.5). □

Theorem 4.1.3. *The variance for the number of immune individuals at time t for the Transient Immunity process* $\mathbf{Y} = (I(t), R(t))$ *with* $\gamma, \beta, \delta > 0$, $\gamma - \beta \neq 0$, *and* $\gamma - \beta \neq \delta$ *and initial conditions* $\mathbf{Y}(0) = (1, 0)$, *is given by*

$$\begin{aligned} Var(R(t)) = {} & \frac{-2\beta\gamma}{(\delta+\gamma-\beta)(\gamma-\beta)} \left[\frac{\beta}{2(\beta-\gamma)} + \frac{\beta}{\delta+\gamma-\beta} + \frac{\gamma}{\gamma-\beta} - \frac{\gamma}{\delta}\right] \\ & - \frac{2\beta\gamma}{(\delta+\gamma-\beta)(\beta e^{(\beta-\gamma)t} - \gamma)} \left[\frac{\beta e^{2(\beta-\gamma)t}}{2(\beta-\gamma)} + \frac{\beta e^{((\beta-\gamma)-\delta)t}}{\delta+(\gamma-\beta)} + \frac{\gamma e^{(\beta-\gamma)t}}{\gamma-\beta} - \frac{\gamma e^{-\delta t}}{\delta}\right] \end{aligned}$$

The proof, found in Appendix A.4, follows by computing the second moment through differential equations, obtained via conditioning on the first event and use of the branching property. Expressions for the cases where $\beta = \gamma$ and $\gamma - \beta = \delta$ can also be found there.

4.1.2 Generating Functions

In addition to the generating function for the Linear BDP, we can generate generating functions related to the Transient Immunity Process. We look at the joint generating function $\mathcal{G}_{IR}(z, w; t) = \mathbb{E}[z^{I(t)} w^{R(t)}]$.

Theorem 4.1.4. *The joint probability generating function for the Transient Immunity Process with the initial conditions* $\mathbf{Y}(0) = (1, 0)$ *is given by*

$$\begin{aligned} \mathcal{G}_{IR}(z, w, t) = {} & \frac{\beta+\gamma}{2\beta} + \frac{\Psi_a'(xe^{-\delta t/2})}{\Psi_a(xe^{-\delta t/2})} \frac{\delta x e^{-\delta t/2}}{2\beta} + \\ & + \Psi_a(xe^{-\delta t/2})^{-2} \left[\varphi(z, x) - \beta \int_0^t \Psi_a(xe^{-\delta\tau/2})^{-2} d\tau\right]^{-1} \end{aligned} \tag{4.1.7}$$

with the notation

$$a = (\gamma - \beta)/\delta \qquad x = \tfrac{2}{\delta}\sqrt{\gamma\beta(1-w)}$$
$$\varphi(z,x) = \left(\left(z - \frac{\beta+\gamma}{2\beta}\right)\Psi_a(x)^2 - \Psi_a'(x)\Psi_a(x)\frac{x\delta}{2\beta}\right)^{-1}$$

and where $\Psi_a(y)$ *is the modified Bessel function of the first kind with parameter* $a \in \mathbb{R}$.

Note that the integral is over a real parameter τ and so the real and imaginary parts are integrated seperately and summed. This follows by constructing the Kolmogorov Forward Equations and using this to determine the PDE which solves to give the p.g.f.

$$\frac{\partial \mathcal{G}_{IR}}{\partial t}(z,w,t) = \left(\beta z^2 - z\beta - z\gamma + w\gamma\right)\frac{\partial \mathcal{G}_{IR}}{\partial z} + \delta(1+w)\frac{\partial \mathcal{G}_{IR}}{\partial w} \tag{4.1.8}$$

with the initial condition $\mathcal{G}_{IR}(s_1, s_2; 0) = s_1$.

A special case of a lengthy calculation discussed in Section 3.2 of Puri [1968] tells us that this solves to obtain the expression (4.1.7). This can be solved numerically, but can't be written in terms of simpler functions. We can use this, however, to obtain the probability generating function of $(I(t), R(t))$ conditional on $\{I(t) > 0\}$ for $\gamma - \beta > 0$, $\delta > 0$.

4.2 Conditioning on Non-extinction

In the rest of this chapter, we will focus on conditioning on two different events, both of which could correspond to a notion of "extinction". In this section, we will consider conditioning on the event $\{I(t) > 0\}$: having a positive number of infectives, as was considered in Section 3.2. In Section 4.3, we will instead condition on the event $\{(I(t), R(t)) \neq (0,0)\}$: there being either infectives or immune individuals remaining in the population. For this section, we assume that all the states $\{(0,r) : r \geq 0\}$ are absorbing, and as such we can identify them all as a single absorbing state. One might use this conditioning to estimate the number of immune individuals in a large population during an epidemic which has not yet gone extinct. This knowledge could be used to inform targets for levels of vaccination in order to improve "herd immunity" in the population, made easier by the natural immunity of recovered

individuals.

In this section, we will determine the expected behaviour and variance of the process conditioned on having a positive number of infectives. To achieve this, we will make use of results from Section 3.1, as well as more direct means. We will also determine the existence and uniqueness of quasi-stationary distributions conditional on there being a positive number of infectives, and some notions of characterising these distributions, including how the marginals of the quasi-stationary distributions relate to QSDs of the Linear BDP.

4.2.1 Existence and Characterisation of QSDs

First, we discuss the existence and uniqueness problem regarding quasi-stationary distributions conditional on there being a positive number of infectives in the population. To prove existence of QSDs, it is sufficient, by Theorem 2.1.17, to prove that the decay parameter α is strictly positive and that asymptotic remoteness holds (Definition 2.1.15). One would like to use Theorem 2.1.14 to use the fact that $\alpha_0 = \alpha$. However, this cannot be applied because absorption can occur from the countably infinite collection of states $\{(1, r) : r \geq 0\}$. Instead we prove existence directly for the limiting conditional distribution as defined in Definition 2.1.4 starting from the state $(1, 0)$. We then find α by use of marginal results which will bound α away from zero. We will make use of properties of quasi-stationary distributions of the Linear BDP to determine properties about the Transient Immunity Process QSDs conditional on $\{I(t) > 0\}$ by considering marginals.

Theorem 4.2.1. *Let $\mathbf{Y}(t) = (I(t), R(t))_{t\geq 0}$ be the Transient Immunity Process with parameters γ, β, $\delta > 0$ with $\gamma > \beta$. Then, the $(1, 0)$-LCD conditional on $\{I > 0\}$ exists, and has the characteristic function:*

$$\chi(s_1, s_2) = (\varphi(0, x) - \varphi(e^{is_1}, x))x^{2a}\frac{(\gamma - \beta)^2}{\beta^2}\frac{\gamma}{\gamma - \beta}\frac{1}{2^{2a}\Gamma(a + 1)^2}$$

with x, a and $\varphi(z, x)$ as defined in (4.2.3).

To do this we will apply Lévy's Continuity Theorem for multivariate processes as given in Theorem 3.9.4 of Durrett [2010].

Theorem 4.2.2. *Let $(\chi_t)_{t\geq 0}$ be a sequence of characteristic functions for a sequence of probability measures μ_t on a common probability space. Suppose that $\chi_t \to \chi$*

pointwise to some function χ as $t \to \infty$, and that χ is continuous at 0. Then χ is the characteristic function for some probability measure μ and μ_t converge in distribution to μ.

This limiting probability measure is the distribution for some random variable, which will in this case be the $(1,0)$-LCD, as defined in Definition 2.1.4. By uniqueness of limits, if the LCD exists, then it will be unique. In general we cannot say about uniqueness of QSDs for the Transient Immunity process.

Proof of Theorem 4.2.1. We find an expression for the conditional characteristic function $\chi_t(s_1, s_2) := \mathbb{E}[e^{i(s_1 I(t) + s_2 R(t))} | I(t) > 0]$ using the expression for the unconditional p.g.f in Theorem 4.1.4. The characteristic function can be seen in terms of the p.g.f. by noting that, where defined, $\mathbb{E}[e^{i(s_1)I(t)+is_2R(t)}] = \mathcal{G}_{IR}(e^{is_1}, e^{is_2}, t)$. We then take the limit as $t \to \infty$ and prove that the limiting characteristic function is continuous at $(s_1, s_2) = (0, 0)$. Then, by Theorem 4.2.2 the limiting characteristic function is a characteristic function for some random variable, which will be precisely the LCD in question. Throughout this proof, we drop subscripts of $\mathbb{P}$ and $\mathbb{E}$ for clarity, all initial conditions are $\mathbf{Y}(0) = (1, 0)$.

Note that, from the derivation in Puri [1968], and by examining the Kolmogorov equations used to derive the p.g.f., we can extend the definition of the p.g.f. to include arguments of the form $z = e^{is_1}$, $w = e^{is_2}$ and obtain the characteristic function χ_t immediately as a change of variables. From the definition of probability generating functions we see that for $z, w \in \mathbb{C}$:

$$\sum_{(j,r)\in\bar{S}} e^{i(s_1 j + s_2 r)} \mathbb{P}[\mathbf{Y}(t) = (j, r), I(t) > 0] = \mathcal{G}_{IR}(e^{is_1}, e^{is_2}, t) - \mathcal{G}_{IR}(0, e^{is_2}, t)$$

where $\bar{S} = \mathbb{N}_0 \times \mathbb{N}_0$. Using this, the conditional characteristic function $\chi_t(s_1, s_2)$ is given by the following:

$$\begin{aligned}
\chi_t(s_1, s_2) &:= \mathbb{E}[e^{i(s_1 I(t) + s_2 R(t))} | I(t) > 0] \\
&= \sum_{(j,r)\in\bar{S}} e^{i(s_1 j + s_2 r)} \mathbb{P}[\mathbf{Y}(t) = (j, r) | I(t) > 0] \\
&= \frac{\chi_t(s_1, s_2) - \mathcal{G}_{IR}(0, e^{is_2}, t)}{\mathbb{P}[I(t) > 0]}.
\end{aligned} \tag{4.2.1}$$

From Theorem 3.2.1 we know that $\mathbb{P}[I(t) > 0] = (\gamma - \beta)/(\gamma e^{(\gamma-\beta)t} - \beta)$. Substituting

in the expression for $\mathcal{G}_{IR}(0, e^{is_2}, t)$ given in (3.2.6), and rearranging we get

$$\chi_t(s_1, s_2) = \Big(\varphi(0, x) - \varphi(e^{is_1}, x)\Big)\, \Psi_a(xe^{-\delta t/2})^2 \Lambda(e^{is_1}, x, t)\Lambda(0, x, t)\frac{\gamma e^{(\gamma-\beta)t} - \beta}{\gamma - \beta} \tag{4.2.2}$$

where, similarly to Theorem 4.1.4, we use the notation for $s_2 \in \mathbb{R}, z, x \in \mathbb{C}, t \geq 0$:

$$\begin{aligned} x = \tfrac{2}{\delta}\sqrt{\gamma\beta(1 - e^{is_2})} \qquad a = (\gamma - \beta)/\delta \\ \varphi(z, x) = \left(\left(z - \frac{\beta + \gamma}{2\beta}\right)\Psi_a(x)^2 - \Psi_a'(x)\Psi_a(x)\frac{x\delta}{2\beta}\right)^{-1} \\ \Lambda(z, x, t) := \frac{\Psi_a(xe^{-\delta t/2})^{-2}}{\varphi(z, x) - \beta\int_0^t \Psi_a(xe^{-\delta\tau/2})^{-2})d\tau} \end{aligned} \tag{4.2.3}$$

To simplify this expression at the limit we note the following properties of the Modified Bessel Function of the first kind, as given in Sections 10.29-10.31 of Clark et al. [2010]. Taking a complex argument $z \in \mathbb{C} \setminus \{a + 0i : a < 0\}$ we have

$$\Psi_a(y) \sim \frac{y^a}{2^a\Gamma(a + 1)} \tag{4.2.4}$$

$$\tfrac{d}{dy}(\Psi_a(y)) = \Psi_{a+1} + ay^{-1}\Psi_a(y) \tag{4.2.5}$$

Simple application of these gives the following:

$$\lim_{y\to 0} \frac{\Psi_a'(y)}{\Psi_a(y)} y = a \qquad a > 0 \tag{4.2.6}$$

$$\lim_{y\to 0} \Psi_a(y)y^{-a} = 2^{-a}\Gamma(a + 1)^{-1} \qquad a > 0 \tag{4.2.7}$$

Using L'Hôpital's rule, chain rule for differentiation and (4.2.6) we get that

$$\lim_{t\to\infty} \Lambda(s_1, x, t) = \frac{-\delta a}{\beta} = \frac{-(\gamma - \beta)}{\beta} \tag{4.2.8}$$

Using (4.2.7) we note that since $\gamma - \beta = \delta a$ we use the product rule and see,

$$\begin{aligned} \lim_{t\to\infty} e^{(\gamma-\beta)t}\Psi_a(xe^{-\delta t/2})^2 &= \lim_{t\to\infty}\left(\frac{\Psi_a(xe^{-\delta t/2})}{\Psi(e^{-\delta t/2})}\right)^2\left(\frac{\Psi_a(e^{-\delta t/2})}{(e^{-\delta t/2})^a}\right)^2 \\ &= \frac{x^2 a}{2^{2a}\Gamma(a + 1)^2} \end{aligned}$$

and hence because $\Psi_a(y) \to 0$ as $y \to 0$,

$$\lim_{t\to\infty} \frac{(\gamma e^{(\gamma-\beta)t} - \beta)\Psi_a(xe^{-\delta t/2})^2}{\gamma - \beta} = \frac{\gamma}{\gamma - \beta} \frac{x^{2a}}{2^{2a}\Gamma(a+1)^2} \tag{4.2.9}$$

We now use the product rule for limits, and substitute (4.2.8) and (4.2.9) into (4.2.2) to obtain the limiting characteristic function $\chi(s_1, s_2) := \lim_{t\to\infty} \chi_t(s_1, s_2)$:

$$\chi(s_1, s_2) = (\varphi(0, x) - \varphi(e^{is_1}, x))x^{2a} \frac{(\gamma - \beta)^2}{\beta^2} \frac{\gamma}{\gamma - \beta} \frac{1}{2^{2a}\Gamma(a+1)^2} \tag{4.2.10}$$

We now need to prove continuity of the characteristic function $\chi(s_1, s_2)$ at $(0, 0)$. Noting that $x = x(s_2)$ is a continuous function of $s_2 \in \mathbb{R}$ and that $x(0) = 0$, we use the composition of continuous functions to allow us to prove that the characteristic function is a continuous function of s_1 and x at $s_1 = 0, x = 0$. More specifically, we observe that

$$\begin{aligned}
&(\varphi(0, x) - \varphi(e^{is_1}, x))x^{2a} = \\
&\frac{-x^{2a}e^{is_1}\Psi_a(x)^2}{\left((-\frac{\gamma+\beta}{2\beta})\Psi_a(x)^2 - \Psi_a'(x)\Psi_a(x)\frac{x\delta}{2\beta}\right)\left((e^{is_1} - \frac{\gamma+\beta}{2\beta})\Psi_a(x)^2 - \Psi_a'(x)\Psi_a(x)\frac{x\delta}{2\beta}\right)} \\
&= \left(\frac{x^a}{\Psi_a(x)}\right)^2 \frac{1}{\left((-\frac{\gamma+\beta}{2\beta}) - \Psi_a'(x)\Psi_a(x)^{-1}\frac{x\delta}{2\beta}\right)\left((e^{is_1} - \frac{\gamma+\beta}{2\beta}) - \Psi_a'(x)\Psi_a(x)^{-1}\frac{x\delta}{2\beta}\right)}
\end{aligned}$$

Using the limits (4.2.6) and (4.2.7) we get

$$\lim_{\substack{s_1\to 0 \\ s_2\to 0}} (\varphi(0, x) - \varphi(e^{is_1}, x))x^{2a} = \frac{\beta^2 2^{2a}\Gamma(a+1)^2}{\gamma(\gamma - \beta)}$$

Substituting into the equation (4.2.10) noting that all the other terms are independent of (x, s_2) we see that $\chi(s_1, s_2) \to 1$ as $(s_1, s_2) \to 0$. The limits involved hold for positive and negative s_1, s_2 so we obtain the limit, and hence we have continuity of the limiting characteristic function. This gives all the conditions for Theorem 4.2.2, so we apply it to obtain a limiting probability distribution, which by construction must be the distribution for the $(1, 0)$-LCD. □

Now we look to characterise this $(1, 0)$-LCD by obtaining its invariance value as defined in Definition 2.1.5. To do this we note the following.

Proposition 4.2.3. *Suppose that* $\mathbf{u} = (u(i, r) : i > 0, r \geq 0)$ *is a proper QSD for*

the Transient Immunity Process $\mathbf{Y}(t) = (I(t), R(t))_{t\geq 0}$ *conditioned on a positive number of infectives, and define the* I-marginal *by* $\mathbf{u}^I = (u^I(i))_{i>0}$

$$u^I(i) = \sum_{r=0}^{\infty} u(i,r) \qquad i \geq 1.$$

Then $\mathbf{u}^I$ *is a QSD for the Linear BDP on* $\mathbb{N}\cup\{0\}$ *conditional on* $\{I > 0\}$*. Moreover, if* $\mathbf{u}$ *is* x*-invariant for the Transient Immunity Process for some* $x > 0$*, then* $\mathbf{u}^I$ *is* x*-invariant for the Linear BDP.*

Proof. To prove the proposition, we let $\mathbf{Y}$ start from an x-invariant QSD $\mathbf{u}$ so that the initial distribution of the I-component is given by $\mathbf{u}^I$. Conditioning on non-absorption we recall Definition 2.1.1 and see that, for $i \geq 1$:

$$\begin{aligned}\mathbb{P}_{\mathbf{u}^I}\left[I(t) = i | T > t\right] &:= \sum_{j=1}^{\infty} \mathbb{P}_j\left[I(t) = i | T > t\right] u^I(j) \\ &= \sum_{j=1}^{\infty}\sum_{s=0}^{\infty} \mathbb{P}_j\left[I(t) = i | T > t\right] u(j,s)\end{aligned}$$

by definition of $\mathbf{u}^I$. Let $\mathcal{F}_t = \sigma(\{\mathbf{Y}(s) : s \leq t\})$ and let $\mathcal{G}_t = \sigma(\{I(s) : s \leq t\})$. Then from construction of the process we note that for any set $A \in \mathcal{F}_t$, we have

$$\mathbb{P}[I(t) \in A | \mathcal{F}_t] = \mathbb{P}[I(t) \in A | \mathcal{G}_t] \tag{4.2.11}$$

We also note the following partition of the state space for each $t \geq 0$, $j > 0$:

$$\{I(t) = j\} = \bigcup_{s=0}^{\infty} \{(I(t), R(t)) = (j,s)\} \tag{4.2.12}$$

Using (4.2.11) and (4.2.12) we obtain

$$\mathbb{P}_{\mathbf{u}^I}\left[I(t) = i | T > t\right] = \sum_{i=1}^{\infty}\sum_{s=0}^{\infty} \mathbb{P}_{(j,s)}\left[I(t) = i | T > t\right] u(j,s)$$

Here, (4.2.11) gives us that $\mathbb{P}_{(j,s)}[\mathbf{Y}(t) = (i,r) \text{ for some } r] = \mathbb{P}_j[I(t) = i]$. Using

(4.2.12) at time t we get:

$$\begin{aligned}\mathbb{P}_{\mathbf{u}^I}\left[I(t)=i|T>t\right] &= \sum_{i=1}^{\infty}\sum_{s=0}^{\infty}\sum_{r=0}^{\infty}\mathbb{P}_{(j,s)}\left[\mathbf{Y}(t)=(i,r)|T>t\right]u(j,s)\\ &= \sum_{r=0}^{\infty}u(i,r)=u^I(i)\end{aligned}$$

which follows because $\mathbf{u}$ is a QSD.

To prove the second part of the proposition, we use Theorem 1 from Artalejo [2012], which states that starting from an x-invariant quasi-stationary distribution $\mathbf{u}(x)$, the time to extinction $T_{\mathbf{u}} \sim \text{Exp}(x)$. As such, since we know both $\mathbf{u}^I$ and $\mathbf{u}$ are QSDs which, by (4.1.2), have the same absorption time $T \sim \text{Exp}(x)$, and since we know $\mathbf{u}$ is x-invariant, $\mathbf{u}^I$ must be x-invariant by Artalejo [2012] too. □

We can also show the following proposition regarding marginals for limiting conditional distributions.

Proposition 4.2.4. *For the Transient Immunity Process $\mathbf{Y} = (I(t), R(t))_{t\geq 0}$, let $\mathbf{v} = (v(i,r) : i > 0, r \geq 0)$ be an initial distribution which gives full weight to S^* and define $\mathbf{v}^I = (v^I(i) : i > 0)$ to be the I-marginal of $\mathbf{v}$. If $\mathbf{u}$ is the $\mathbf{v}$-LCD for the Transient Immunity Process, then $\mathbf{u}^I$ is the $\mathbf{v}^I$-LCD for the Linear BDP.*

Proof. Using the notation from Equation (2.1.1) we show that $\mathbf{u}^I$ is indeed the $\mathbf{v}^I$-LCD. As in Proposition 4.2.3, we see that

$$\begin{aligned}\lim_{t\to\infty}\mathbb{P}_{\mathbf{v}^I}[I(t)=i|I(t)>0] &= \lim_{t\to\infty}\sum_{j=1}^{\infty}v^I(j)\mathbb{P}_i[I(t)=i|I(t)>0]\\ &= \lim_{t\to\infty}\sum_{j=1}^{\infty}\sum_{s=0}^{\infty}v(j,s)\mathbb{P}_j[I(t)=i|I(t)>0] =: \mathcal{A}\end{aligned}$$

Using the partition in (4.2.12) and that $\mathbb{P}_{(j,s)}[I(t)=i] = \mathbb{P}_j[I(t)=i]$ as in Proposition 4.2.3 we get

$$\mathcal{A} = \lim_{t\to\infty}\sum_{j=1}^{\infty}\sum_{s=0}^{\infty}\sum_{r=0}^{\infty}v(j,s)\mathbb{P}_{(j,s)}[\mathbf{Y}(t)=(i,r)|I(t)>0]$$

We sum over r, and since we are under a proper probability distribution, we can

apply the Dominated Convergence Theorem to exchange limits and sums to obtain

$$\begin{aligned}\mathcal{A} &= \sum_{r=0}^{\infty} \lim_{t\to\infty} \sum_{j=1}^{\infty}\sum_{s=0}^{\infty} v(j,s)\mathbb{P}_{(j,s)}[I(t)=i, R(t)=r | I(t)>0] \\ &= \sum_{r=0}^{\infty} \lim_{t\to\infty} \mathbb{P}_{\mathbf{v}}[\mathbf{Y}(t)=(i,r)|I(t)>0]\end{aligned}$$

Since $\mathbf{u}$ is the $\mathbf{v}$-LCD we see that $\mathcal{A} = \sum_{r=0}^{\infty} u(i,r) = u^I(i)$, and so $\mathbf{u}^I$ is indeed the $\mathbf{v}^I$-LCD of the Linear BDP. □

Note that the above result also tells us that the invariance values of the (j,s)-LCD and the j-LCD match.

With this in hand we now obtain the invariance value for the $(1,0)$-LCD and hence the decay parameter of $\mathbf{Y}(t)$ conditional on $\{I>0\}$.

Theorem 4.2.5. *Let $\mathbf{Y}(t) = (I(t), R(t))_{t\geq 0}$ be the Transient Immunity Process with parameters γ, β, $\delta > 0$ with $\gamma > \beta$. Conditional on $\{I>0\}$, the $(1,0)$-LCD is $(\gamma-\beta)$-invariant, and hence the decay parameter for the Transient Immunity Process conditional on $\{I>0\}$ is $\alpha = \gamma - \beta$.*

Proof. We first note that, by Proposition 4.2.4, the marginal of the $(1,0)$-LCD is the 1-LCD for the Linear BDP. Using Proposition 4.2.3, and the fact that the 1-LCD is $(\gamma-\beta)$-invariant by Proposition 3.2.4, we see that the $(1,0)$-LCD must also be $(\gamma-\beta)$-invariant. Additionally, we note that for the Transient Immunity Process, α_0 as defined in Definition 2.1.7 is $\alpha_0 = \gamma - \beta$ since the time to extinction only depends on the behaviour of $I(t)$, and so is the same as for the value in the Linear BDP case. Finally we note that the limit in Definition 2.1.4 for the LCD exists. Since we know $\alpha \geq \alpha_0 > 0$ we apply Theorem 13 of van Doorn and Pollett [2013] and have that $\alpha = \alpha_0 = \gamma - \beta$ is the decay parameter for the Transient Immunity Process conditional on $\{I>0\}$, and the result is shown. □

As an immediate consequence of Proposition 4.2.3, we observe that for an x-invariant QSD $\mathbf{u}(x)$ for the Transient Immunity Process, we must have $\sum iu^I(i) = \infty$ since this is the case for the Linear BDP. For the expected number of immune individuals, finite expectation is less immediate, but in the case of the α-invariant distribution, the expected number of immune individuals under $\mathbf{u}(\alpha)$ can be computed directly via the $(1,0)$-LCD for example. For $x < \alpha$ we note that since we expect an infinite

number of infectives under the QSD, we would expect an infinite number of individuals coming immune in any period $[t, t+h)$, suggesting an infinite number of immune individuals also.

4.2.2 Expected Behaviour Under Conditioning

We now look into the expected temporal behaviour of the Transient Immunity Process conditional on having a positive number of infectives. This can then be extended to obtain properties of the **v**-LCD starting from a distribution **v** of finite support.

In this section, unless otherwise stated, we assume the initial conditions of the process to be $I(0) = 1, R(0) = 0$. First of all we recall that, from Equation (3.2.3) in Theorem 3.2.1 that

$$P_A(t) := \mathbb{P}_{(1,0)}[I(t) > 0] = \frac{\gamma - \beta}{\gamma - \beta e^{(\beta-\gamma)t}} \tag{4.2.13}$$

The use of the notation $P_A(t)$ will continue in the rest of this chapter. We use this to derive the expectation of the number of infectives conditioned on non-extinction at time $t \geq 0$. Ideally, we would obtain these directly from $\chi(z, w, t)$, the conditional characteristic function obtained in Theorem 4.2.1. However, taking further derivatives in the functions which depends on Modified Bessel functions is difficult. Instead, we construct the differential equations for the expectation and variance directly.

The expected behaviour of the number of infectives for the Transient Immunity Process $\mathbf{Y}(t) = (I(t), R(t))_{t\geq 0}$ with initial conditions $\mathbf{Y}(0) = (1, 0)$ conditioning on a positive number of infectives at time t is given by

$$\mathbb{E}_{(1,0)}[I(t)|I(t) > 0] = \frac{\mathbb{E}_{(1,0)}[I(t)]}{\mathbb{P}_{(1,0)}[I(t) > 0]} = \frac{\beta e^{(\beta-\gamma)t} - \gamma}{\beta - \gamma} \tag{4.2.14}$$

This follows directly from the definition and through Bayes' rule.

Secondly, we look at the expected number of immune individuals under the same conditioning.

Theorem 4.2.6. *The expected number of immune individuals for the Transient Immunity Process* $\mathbf{Y} = (I(t), R(t))_{t\geq 0}$ *conditioned on the event* $\{I(t) > 0\}$ *with*

$\beta, \gamma, \delta > 0$ and $\gamma - \beta \neq \delta$, $\mathbb{E}_{(1,0)}[R(t)|I(t) > 0]$, is given by

$$\mathbb{E}_{(1,0)}[R(t)|I(t) > 0] = \frac{2\beta\gamma}{(\delta + \beta - \gamma)(\beta e^{(\beta-\gamma)t} - \gamma)} \left[\frac{\beta}{2(\beta - \gamma)} (e^{2(\beta-\gamma)t} - 1) \right.$$
$$\left. + \frac{\beta}{(\gamma - \beta) + \delta} (e^{((\beta-\gamma)-\delta)t} - 1) + \frac{\gamma}{\gamma - \beta} (e^{(\beta-\gamma)t} - 1) - \frac{\gamma}{\delta} (e^{-\delta t} - 1) \right]$$

To prove Theorem 4.2.6, we condition on the first event in the time interval $(0, h)$. In order to evaluate this and form a differential equation, we use the following lemma.

Lemma 4.2.7. *For the Transient Immunity Process $\mathbf{Y} = (I(t), R(t))_{t \geq 0}$, we have that, conditioning on the event $\{I(t) > 0\}$:*

$$\mathbb{E}_{(2,0)}[R(t)|I(t) > 0] = \frac{2(1 - P_A(t))}{2 - P_A(t)} \mathbb{E}_{(1,0)}[R(t)|I(t) > 0] + \frac{2}{2 - P_A(t)} \mathbb{E}_{(1,0)}[R(t)]$$

Proof. To decompose $\mathbb{E}_{(2,0)}[R(t)|I(t) > 0]$ into terms we have expressions for, we decompose into two independent numbered processes $\mathbf{Y}^{(1)}(t), \mathbf{Y}^{(2)}(t)$ such that $\mathbf{Y}(t) = \mathbf{Y}^{(1)}(t) + \mathbf{Y}^{(2)}(t)$ with corresponding absorption times $T^{(1)}, T^{(2)}$. By construction,

$$T := \inf\{t : I(t) = 0\} = \max(T^{(1)}, T^{(2)})$$

In this case,we note that the event $\{I(t) > 0\} = \{T > t\} = \{T^{(1)} > t \text{ or } T^{(2)} > t\}$, and this can be decomposed into the disjoint events

$$\{T^{(1)} > t, T^{(2)} \leq t\} \cup \{T^{(1)} \leq t, T^{(2)} > t\} \cup \{T^{(1)} > t, T^{(2)} > t\}$$

Since $Y^{(1)}(t)$ and $Y^{(2)}(t)$ are identically distributed with the same distribution as $Y(t)$ with a single initial infective, we can obtain the following probabilities for the above events:

$$\mathbb{P}_{(1,0)}[T^{(1)} > t, T^{(2)} \leq t | T > t] = \frac{P_A(t)(1 - P_A(t))}{2P_A(t) - P_A(t)^2} = \frac{1 - P_A(t)}{2 - P_A(t)}$$
$$\mathbb{P}_{(1,0)}[T^{(1)} \leq t, T^{(2)} > t | T > t] = \frac{P_A(t)(1 - P_A(t))}{2P_A(t) - P_A(t)^2} = \frac{1 - P_A(t)}{2 - P_A(t)}$$
$$\mathbb{P}_{(1,0)}[T^{(1)} > t, T^{(2)} > t | T > t] = \frac{P_A(t)^2}{2P_A(t) - P_A(t)^2} = \frac{P_A(t)}{2 - P_A(t)}$$

Using these we decompose $\mathbb{E}_{(2,0)}[R_t|I(t) > 0]$ as follows. The initial conditions for $\mathbf{Y}^{(1)}(t)$ and $\mathbf{Y}^{(2)}(t)$ are always a single infective unless otherwise stated in the

subscripts on $\mathbb{E}$.

$$\begin{aligned}\mathcal{A} &:= \mathbb{E}_{(2,0)}[R(t)|I(t)>0] \\ &= \mathbb{E}[R^{(1)}(t)+R^{(2)}(t)|T^{(1)}>t, T^{(2)}\leq t]\left(\frac{1-P_A(t)}{2-P_A(t)}\right) \\ &\quad + \mathbb{E}[R^{(1)}(t)+R^{(2)}(t)|T^{(1)}\leq t, T^{(2)}>t]\left(\frac{1-P_A(t)}{2-P_A(t)}\right) \\ &\quad + \mathbb{E}[R^{(1)}(t)+R^{(2)}(t)|T^{(1)}>t, T^{(2)}>t]\left(\frac{P_A(t)}{2-P_A(t)}\right) \end{aligned} \tag{4.2.15}$$

By the i.i.d. nature of $Y^{(1)}$ and $Y^{(2)}$ we get

$$= \mathbb{E}[R(t)|I(t)>0]\left(\frac{2}{2-P_A(t)}\right) + 2\mathbb{E}[R(t)|I(t)=0]\left(\frac{1-P_A(t)}{2-P_A(t)}\right) \tag{4.2.16}$$

Using the definition of expectation and conditional probability we also observe that

$$\begin{aligned}\mathbb{E}[R(t)|I(t)=0] &= \sum_{r=0}^{\infty}\frac{r\mathbb{P}[R(t)=r, I(t)=0]}{\mathbb{P}[I(t)=0]} \\ &= \sum_{r=0}^{\infty}\frac{r(\mathbb{P}[R(t)=r]-\mathbb{P}[R(t)=r, I(t)>0])}{1-\mathbb{P}[I(t)>0]} \\ &= \frac{\mathbb{E}[R(t)]}{1-P_A(t)} - \mathbb{E}[R(t)|I(t)>0]\frac{P_A(t)}{1-P_A(t)} \end{aligned} \tag{4.2.17}$$

Substituting equation (4.2.17) into equation (4.2.15) we get

$$\begin{aligned}\mathcal{A} &= \mathbb{E}[R(t)|I(t)>0]\frac{2(1-P_A(t))}{2-P_A(t)} + \frac{2\left(\mathbb{E}[R(t)]-\mathbb{E}[R(t)|I(t)>0]P_A(t)\right)}{2-P_A(t)} \\ &= \frac{2(1-P_A(t))}{2-P_A(t)}\mathbb{E}[R(t)|I(t)>0] + \frac{2}{2-P_A(t)}\mathbb{E}[R(t)] \end{aligned} \tag{4.2.18}$$

as required to prove the lemma. □

We also use the following result which looks at the probability of non-absorption from two individuals. Since this also holds for the Linear BDP, and only concerns the I-component, we only consider the I-component in this lemma.

Lemma 4.2.8. *For $P_A(t)$ defined in* (4.2.13) *we have that*

$$\mathbb{P}_2[I(t)>0] = P_A(t)(2-P_A(t))$$

Proof. Using the expression for the absorption probability from $I(0)=i_0$ in 3.2.3

we note that $\mathbb{P}_2[I(t) = 0] = \mathbb{P}_1[I(t) = 0]^2$. Rearranging this and considering the complement we get

$$\begin{aligned}\mathbb{P}_2[I(t) > 0] &= 1 - \mathbb{P}_2[I(t) > 0] \\ &= 1 - (1 - \mathbb{P}_1[I(t) > 0])^2 \\ &= 2\mathbb{P}_1[I(t) > 0] - \mathbb{P}_1[I(t) > 0]^2 \\ &= P_A(t)(2 - P_A(t))\end{aligned}$$

and the result is achieved. □

The proof of Theorem 4.2.6 uses Lemma 4.2.7 and the branching property to construct the differential equation which is solved to obtain the result. See Appendix A.4 for the full proof.

We can also determine the variance of the process conditional on a positive number of infectives up to a given time $t \geq 0$.

Theorem 4.2.9. *The variance for the number of infectives of the Transient Immunity conditional on a positive number of infectives is given by*

$$Var(I(t)|I(t) > 0) = \frac{(2\beta e^{(\beta-\gamma)t} - (\beta+\gamma))(\beta e^{(\beta-\gamma)t} - \gamma)}{(\beta-\gamma)^2} - \frac{(\beta e^{(\beta-\gamma)t} - \gamma)^2}{(\beta-\gamma)^2}$$

This follows from the definition of conditional expectation $\mathbb{E}[I(t)|I(t) > 0]$, and then the application of the second moment $\mathbb{E}[I(t)^2]$ which was found in Theorem 4.1.2 to find $\mathbb{E}[I(t)^2|I(t) > 0]$. For the full proof see Appendix A.4.

Note that, for $\beta < \gamma$ and taking $t \to \infty$, we see that the mean and variance for the number of infectives, conditional on $\{I(t) > 0\}$, converge to coincide with the mean and variance for the 1-LCD of the Linear BDP; a Geometric distribution with parameter $1 - \beta/\gamma$. If $\beta \geq \gamma$ the mean and variance converge to infinity instead.

Theorem 4.2.10. *The variance of the number of immune individuals of the Transient Immunity Process conditional on having a positive number of infectives is given by* $Var(R(t)|I(t) > 0)$ *which satisfies*

$$Var(R(t)|I(t) > 0) = z(t) - \mathbb{E}[R(t)|I(t) > 0]^2 \tag{4.2.19}$$

where $z(t) = \mathbb{E}[R(t)^2|I(t) > 0]$ *is the solution to an equation of the form*

$$z'(t) = z(t)\left[\frac{(4\beta-\gamma)(\gamma-\beta)}{\gamma-\beta e^{(\beta-\gamma)t}} + (\beta+\gamma)\right] + \sum_{j=0}^{12} A_j e^{\lambda_j t} + A_{13}$$

for two sequences of known constants $\{A_0, \ldots, A_{13}\}$ *and* $\{\lambda_0, \ldots, \lambda_{12}\}$.

In order to construct the differential equation we first give the following result.

Lemma 4.2.11. *For the Transient Immunity Process* $\mathbf{Y}(t) = (I(t), R(t))$ *we have*

$$\begin{aligned}\mathbb{E}_{(2,0)}[R(t)^2|I(t)>0] = &\\ &\frac{2\left[(1-2P_A(t))\mathbb{E}[R(t)^2|I(t)>0] - 2P_A(t)\mathbb{E}[R(t)|I(t)>0]^2\right.}{2-P_A(t)}\\ &+\frac{\left.4\mathbb{E}[R(t)]\mathbb{E}[R(t)|I(t)>0] + 2\mathbb{E}[R(t)^2]\right]}{2-P_A(t)}\end{aligned}$$

The proof of this lemma follows in a similar fashion to that of Lemma 4.2.7. Theorem 4.2.10 is then proven using the same methods as Theorem 4.1.2, making use of $\mathbb{E}[R(t)]$ from Theorem 4.1.1, $\mathbb{E}[R(t)^2]$ from Theorem 4.1.3, $\mathbb{E}[R(t)|I(t) > 0]$ from Theorem 4.2.6 and Lemma 4.2.11 to derive the differential equation

$$\begin{aligned}z'(t) = z(t)&\left[4\beta P_A(t) + (\beta+\gamma) - \frac{P_A'(t)}{P_A(t)}\right] + 2\beta P_A(t)\mathbb{E}[R(t)|I(t)>0]^2\\ &+ 4\beta\mathbb{E}[R(t)|I(t)>0]\mathbb{E}[R(t)] + 2\beta\mathbb{E}[R(t)^2]\end{aligned}$$

All the expressions above are known and can be substituted in and then the differential equation of the form stated in the theorem can then be solved through standard techniques. Both the proof of Lemma 4.2.11 and of Theorem 4.2.10 can be found in Appendix A.4.

Limiting Mean Behaviour

Now we have 4.2.14 and Theorem 4.2.6, we can use these to characterise the $(1,0)$-LCD of the Transient Immunity process conditioned on $\{I(t) > 0\}$. We already have existence and uniqueness of the $(1,0)$-LCD in Theorem 4.2.1, but we do not have analytic descriptions. With our explicit expressions for $\mathbb{E}[I(t)|I(t) > 0]$ and $\mathbb{E}[R(t)|I(t) > 0]$ we know that the mean of the LCD coincides with the limits of these expressions.

Theorem 4.2.12. *Starting from a single initial infective $I(0) = 1, R(0) = 0$, and conditional on the event $\{I(t) \neq 0\}$, the $(1,0)$-LCD of the Transient Immunity process $(I(t), R(t))$ has mean value given by*

$$\lim_{t\to\infty} \mathbb{E}[I(t)|I(t) > 0] = \frac{\gamma}{\gamma - \beta} \tag{4.2.20}$$

$$\lim_{t\to\infty} \mathbb{E}[R(t)|I(t) > 0] = \frac{2\gamma}{(\beta - \gamma) + \delta}\left(\frac{\beta - 2\gamma}{2(\gamma - \beta)} - \frac{\beta}{(\gamma - \beta) + \delta} + \frac{\gamma}{\delta}\right) \tag{4.2.21}$$

These both follow simply from the expressions in 4.2.14 and Theorem 4.2.6.

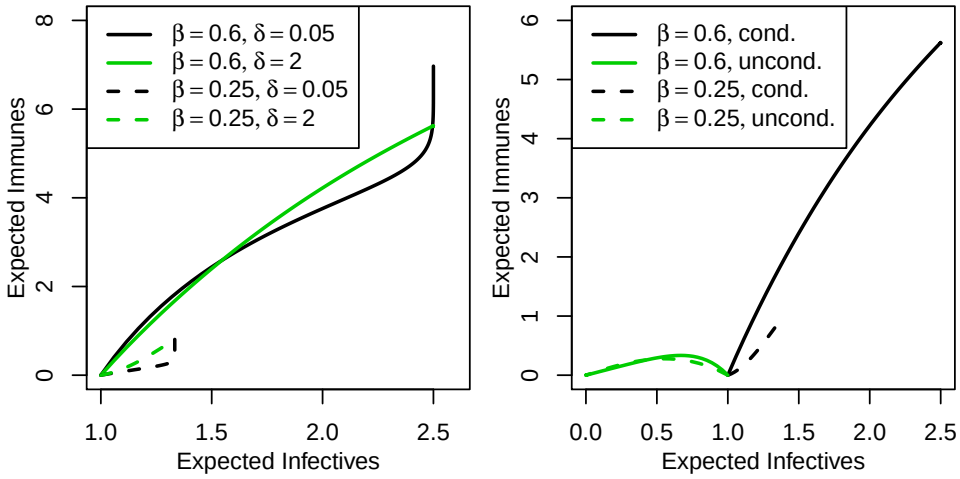
Remark 4.2.13. At this point we make the following remarks:

- We see that Equation (4.2.20) matches the known properties of the 1-LCD for the Linear BDP as described in Proposition 3.2.4. If one takes the limit as $t \to \infty$ of the variance $\mathrm{Var}(I(t)|I(t) > 0)$ given in Theorem 4.2.9, we also obtain that this limit equals the variance of the 1-LCD.
- Furthermore, we have that the expected number of immune individuals under the $(1,0)$-LCD is finite for all values of γ, β, δ with $(\beta - \gamma) < 0$, $\delta > 0$. However, one can see that as $(\gamma - \beta) \to 0$ or $\delta \to 0$, we get $\mathbb{E}_{(1,0)}[R(t)|I(t) > 0] \to \infty$, and as $(\gamma - \beta) \to \infty$ or $\delta \to \infty$ we get $\mathbb{E}_{(1,0)}[R(t)|I(t) > 0] \to 0$.

In Figure 4.2.1a, we see the evolution of the expected behaviour of the Transient Immunity process given $I(t) > 0$. For fixed δ, as β increases for $\gamma = 1$, the limit as $t \to \infty$ increases in both components, and for fixed $\beta - \gamma$, only the R component increases, as expected. Compare to the unconditioned case in Figure 4.2.1b in the case where $\delta = 2$; the conditioning quickly makes a difference in behaviour.

It should be noted here that in Figure 4.2.1a, the curves have in fact converged to these values, these differential equations were solved to timepoints well past the point at which they appear to reach their limit points. Note however, the solutions only reach the values of the expectations of the corresponding QSDs at the limit as $t \to \infty$. Note the unconditioned behaviour converges to zero in all cases, since $\beta < \gamma$.

All the results in Section 4.2.2 pertaining to the expected behaviour (4.2.14, Theorems 4.2.6, 4.2.9, 4.2.10) do in fact hold in the supercritical case where $\gamma - \beta < 0$, but this supercritical case leads to non-certain absorption and so there does not exist a QSD, and hence no limiting conditional distribution in this case. In this



(a) Expected values conditional on $I(t) > 0$ as β, δ vary with $\gamma = 1$

(b) Expected values as β varies, $\delta = 2$, $\gamma = 1$ with and without conditioning

Figure 4.2.1: Expected values of $\mathbf{Y}(t)$ with and without conditioning on $I(t) > 0$. Numerical calculations for conditional behaviour solved on the interval $t \in [0, 400]$.

case, the expected number of infectives and hence immune individuals converge to infinity as $t \to \infty$.

4.3 Conditioning on Remaining Immunity

In this section we consider an alternative event on which to condition the QSD: there still being either infectives or immune individuals still remaining in the population. This means that the Transient Immunity Process has a single absorbing state $\{(i,r) = (0,0)\}$. We note, however, that under this notion of absorption, the process is no longer irreducible. In addition to the countably infinite communicating class S^* defined in (4.1.1) we also have the infinite sequence of single-state communicating classes $S_r = \{(0,r)\}$ for $r > 0$. As such we can no longer immediately use Theorems 2.1.16 and 2.1.17, as the state space is neither finite nor a single countable communicating class. However, we will still be able to make some conclusions about the existence and uniqueness of quasi-stationary distributions under this conditioning. It should be noted at this point that it is in fact impossible for the process to jump straight from $\{(i,r) : i > 0, r \geq 0\}$ to $(0,0)$ since the last infective must first become immune before losing that immunity.

Letting $B(t)$ denote the event $\{\mathbf{Y}(t) \neq (0,0)\}$, we will obtain descriptions of the probability of absorption after time t. We will also discuss the existence and uniqueness problem and characterise the $\mathbf{v}$-LCDs for $\mathbf{v}$ with finite support. We will conclude with some discussion and illustration of the expected behaviour of $\mathbf{Y}(t)$ under this conditioning. We begin with the probability of extinction.

Proposition 4.3.1. *The probability of non-absorption* $P_B(t) := \mathbb{P}_{(1,0)}[\mathbf{Y}(t) \neq (0,0)]$ *with initial condition* $P_B(0) = 1$ *solves the differential equation:*

$$P_B'(t) = -\beta P_B(t)^2 + (\beta - \gamma)P_B(t) + \gamma e^{-\delta t}. \tag{4.3.1}$$

Proof. Let $B(t) = \{\mathbf{Y}(t) \neq (0,0)\}$ be as above. Unless otherwise specified, we suppress conditioning on the event $\{\mathbf{Y}(0) = (1,0)\}$ in all relevant notation.

Using the backward Kolmogorov equations for the Transient Immunity process with initial condition $\{\mathbf{Y}(0) = (1,0)\}$ we obtain the expression

$$P_B'(t) = \beta \mathbb{P}_{(2,0)}[B(t)] + \gamma \mathbb{P}_{(0,1)}[B(t)] - (\beta + \gamma)P_B(t) \tag{4.3.2}$$

Using a similar argument to that of Theorem 4.2.9, we also see that

$$\mathbb{P}_{(2,0)}[B(t)] = 2P_B(t) - P_B(t)^2$$

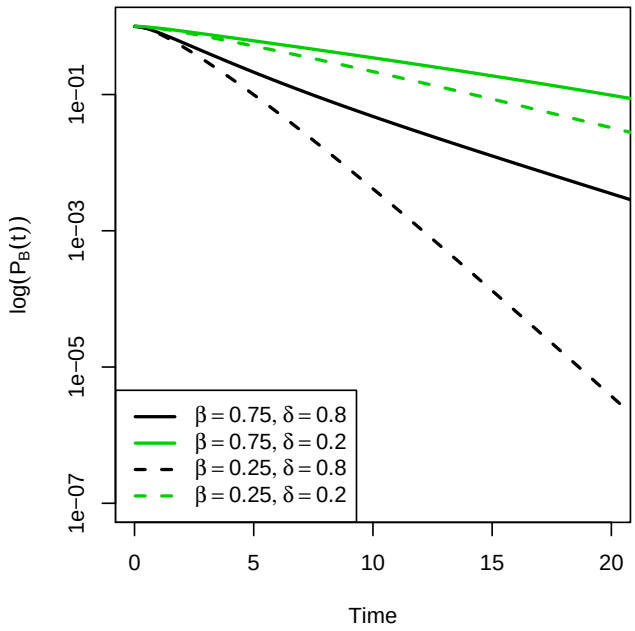


Figure 4.3.1: Probability of extinction of Transient Immunity Process conditional on $\{\mathbf{Y}(t) \neq (0,0)\}$ for varying β, δ with $\gamma = 1$

Since there is only one possible event (loss of immunity from one individual) which can happen starting from $\mathbf{Y}(0) = (0,1)$ we see that

$$\mathbb{P}_{(0,1)}[B(t)] = \mathbb{P}[\text{Exp}(\delta) > t] = e^{-\delta t}$$

Substituting these into (4.3.2) we obtain the result. □

This equation can be solved, but only in terms of a lengthy expansion of modified Bessel functions. Since we only describe $P_B(t)$ using a differential equation description, we will note some properties of the limiting behaviour. Figure 4.3.1 gives some examples of the evolution of $P_B(t)$ for different values of $\beta, \gamma, \delta > 0$. We plot the figure on a logarithmic scale since $P_B(t)$ converges to zero, but all values of the parameters show exponential decay of $P_B(t)$.

4.3.1 Existence and Characterisation of QSDs

In order to determine the existence and uniqueness of QSDs conditional on $\{\mathbf{Y}(t) \neq (0,0)\}$, we note that the state space S is neither finite nor irreducible, which are the main points of focus in the literature of this field. To this end we will discuss

existence of QSDs and moreover determine the existence of QSDs with positive weight on $S^* = \{(i,r) : i > 0, r \geq 0\}$. These will also be illustrated using simulation methods which we will explain and develop in Chapter 5.

We will prove the following theorem which outline the existence of QSDs.

Theorem 4.3.2. *For the Transient Immunity Process* $\mathbf{Y}(t)$ *conditional on* $\{\mathbf{Y} \neq (0,0)\}$ *we observe that:*

- *For all values of* β, γ, $\delta > 0$ *such that* $\gamma - \beta > 0$ *we have that* $\mathbb{1}_{(0,1)}(i,r)$, *the point mass on* $\{(0,1)\}$ *is a* δ*-invariant QSD.*
- *Additionally, for each* $x \in (0, \delta]$ *we have that* $\mathbf{u}(x)$ *defined by*

$$\begin{aligned} u(0,1) &= x\delta^{-1} \\ u(0,n) &= \frac{x}{\delta n!} \prod_{i=2}^{n} ((i-1) - x/\delta) \qquad n \geq 2 \end{aligned} \tag{4.3.3}$$

  is an x*-invariant QSD.*

Proof. This theorem follows since if the initial condition of the Transient Immunity Process gives full weight to the region $\{(0,r) : r > 0\}$ then the R-component behaves identically to the linear pure death process investigated in Theorem 3.3.2. As such, all the QSDs stated there also hold here by setting the number of infectives to be constantly equal to zero, and the number of immune individuals set to the number of individuals in the death process under the QSD. □

We now consider the more interesting case of existence of QSDs with weight on $S^* = \{(i,r) : i \geq 1, r \geq 0\}$. In order to do this, we focus on the $(1,0)$-LCD, but this could be extended to any (i,r)-LCD for $(i,r) \in S^*$, and in fact any $\mathbf{v}$ with finite support contained in S^*. In order to do this we need the following technical result.

Proposition 4.3.3. *Let* Φ *be defined for the Transient Immunity Process* $\mathbf{Y} = (I(t), R(t))_{t \geq 0}$ *with initial conditions* $\mathbf{Y}(0) = (1,0)$ *by*

$$\Phi := \lim_{t \to \infty} \frac{\mathbb{P}[I(t) > 0]}{\mathbb{P}[\mathbf{Y}(t) \neq (0,0)]}$$

Then $\Phi > 0$ *precisely when* β, γ, $\delta > 0$ *and* $0 < \gamma - \beta < \delta$.

Proof of Proposition 4.3.3. Let $P_B(t)$ and $P_A(t)$ be as defined at the start of Sections 4.2 and 4.3, and let $u(t)$ be such that $P_B(t) = u(t)P_A(t)$. Then using the differential equations for P_A and P_B we can obtain a differential equation for $u(t)$. Using the product rule and noting that

$$P_A'(t) = -\beta P_A(t)^2 + (\beta - \gamma)P_A(t)$$

we can find that, substituting into (4.3.1) we obtain

$$u'(t)P_A(t) + u(t)P_A'(t) = -\beta u(t)^2 P_A(t)^2 + (\beta - \gamma)u(t)P_A(t) + \gamma e^{-\delta t}$$

Which can be rearranged to get

$$u'(t) = \beta P_A(t)(u(t) - u(t)^2) + \frac{\gamma e^{-\delta t}}{P_A(t)}$$

Substituting in the expression for $P_A(t)$ in equation (4.2.13) we get

$$u'(t) = \frac{\beta(\beta - \gamma)}{\beta - \gamma e^{(\gamma-\beta)t}}(u(t) - u(t)^2) + \frac{\gamma\beta e^{-\delta t} - \gamma^2 e^{((\gamma-\beta)-\delta)t)}}{\beta - \gamma} \tag{4.3.4}$$

We consider the three cases $\gamma - \beta < \delta$, $\gamma - \beta = \delta$ and $\gamma - \beta > \delta$. In the first case, the second term on the right of (4.3.4) decays exponentially to zero. Since $u(t) \geq 1$ for all $t \geq 0$, the first term is negative, so we must have that $u'(t)$ is eventually negative and hence $u(t) \to C$ for some constant $1 \leq C < \infty$.

In the second case, we note that $u'(t) \leq K$ for all $t \geq 0$ with $K = \gamma^2/(\gamma - \beta) > 0$ since all other contributions to (4.3.4) are negative. As such, we must have $u(t) \leq u(0) + Kt$ and so since the exponential decay $\frac{\beta(\beta-\gamma)}{\beta-\gamma e^{(\gamma-\beta)t}}$ in the first term is faster than the polynomial growth of $|Kt - K^2t^2|$, we must have that the first term decays to zero, and hence $u'(t) \to K$. Since $K > 0$ then we can conclude $u(t)$ tends to ∞ as $t \to \infty$.

In the final case, let $P_B(t, \delta) = \mathbb{P}[\mathbf{Y}(t) \neq (0, 0)]$ given the loss of immunity rate is equal to δ. By definition of the Transient Immunity Process, we observe that reducing δ reduces the probability of entering $(0, 0)$ for fixed t, and so $P_B(t, \delta) \geq P_B(t, \delta - \varepsilon)$ for all $t \geq 0, \varepsilon > 0$. Since $P_A(t)$ is independent of δ, we conclude that, if $u(t, \delta)$ is the value of u given a choice of δ, then $u(t, \delta - \varepsilon) \geq u(t, \delta)$. Therefore, since $u(t, \gamma - \beta) \to \infty$, so must $u(t, \delta)$ for $\delta < \gamma - \beta$.

The theorem then follows by noting that $\Phi = \lim_{t\to\infty} 1/u(t)$. □

Now we prove existence of the $(1,0)$-LCD.

Theorem 4.3.4. *For the Transient Immunity Process process with $\beta, \gamma, \delta > 0$ and $\beta < \gamma$ we have existence and uniqueness of the $(1,0)$-LCD conditional on $\{(I(t), R(t)) \neq (0,0)\}$.*

Proof. Let $\mathbf{u}^A = (u^A(i,r))_{(i,r)\in S^*}$ be the $(1,0)$-LCD for the Transient Immunity Process conditional on $\{I > 0\}$, with S^* defined in (4.1.1). Fix $(i,r) \in S^*$. Then we have that, by the definition of conditional probability,

$$\begin{aligned}
&\lim_{t\to\infty} \mathbb{P}_{(1,0)}[\mathbf{Y}(t) = (i,r) | \mathbf{Y}(t) \neq (0,0)] \\
&\qquad = \lim_{t\to\infty} \mathbb{P}_{(1,0)}[\mathbf{Y}(t) = (i,r) | I(t) > 0] \frac{\mathbb{P}_{(1,0)}[I(t) > 0]}{\mathbb{P}_{(1,0)}[\mathbf{Y}(t) \neq (0,0)]} \\
&\qquad = \lim_{t\to\infty} \mathbb{P}_{(1,0)}[\mathbf{Y}(t) = (i,r) | I(t) > 0] \lim_{t\to\infty} \frac{\mathbb{P}_{(1,0)}[I(t) > 0]}{\mathbb{P}_{(1,0)}[\mathbf{Y}(t) \neq (0,0)]}
\end{aligned}$$

The product rule holds since, by Proposition 4.3.3, the fraction term is bounded and converges to Φ and by Theorem 4.2.1 $\lim_{t\to\infty} \mathbb{P}_{(1,0)}[\mathbf{Y}(t) = (i,r) | I(t) > 0]$ converges to $\mathbf{u}^A(i,r)$. Given these limits exist, we need only consider the values of the LCD for states $(0,r)$ with $r > 0$. Such an LCD is also a QSD and therefore is an x-invariant QSD for some value $x > 0$. Using this x-invariance, we can iteratively obtain the values $u(0,r)$ by noting that

$$u(1,0)\beta + 2\delta u(0,2) - \delta u(0,1) = -xu(0,1)$$

and subsequent values can be found iteratively, and we use the value of Φ to normalise these values, and obtain the $(1,0)$-LCD in this case. Note this holds even if $\Phi = 0$. Uniqueness follows by uniqueness of the limits and of solutions to the x-invariance equation. □

Now we arrive at the main theorem of this section, which follows using the propositions above.

Theorem 4.3.5. *For the Transient Immunity Process $\mathbf{Y}(t)$ conditional on $\mathbf{Y} \neq (0,0)$ with parameters $\gamma, \beta, \delta > 0$ with $\gamma > \beta$, let $\alpha^* = \gamma - \beta$, and let $\mathbf{u}$ be the $(1,0)$-LCD for the Transient Immunity Process conditional on $\{\mathbf{Y}(t) \neq (0,0)\}$. Then the following hold:*

- *If $\alpha^* = \gamma - \beta < \delta$ then the $(1,0)$-LCD gives positive weight to S^*, where S^* is*

as in (4.1.1).

- *If* $\alpha^* \geq \delta$ *then the* $(1,0)$*-LCD has no weight on* S^**, and hence is precisely the unit mass on* $(0,1)$.
- *The* $(1,0)$*-LCD is* $\min(\delta, \alpha^*)$*-invariant.*

Proof. The first point follows directly from Proposition 4.3.3. If $\Phi = \mathbf{u}(S^*) = 0$, then we know all the mass must be on $S \backslash S^* = \{(0,r) : r > 0\}$. We note that starting from a single point that the distribution for the number of infectives maintains a finite mean for all time. If the LCD were to have a finite mean number of immune individuals, then restarting from such a distribution and conditioning on $\{I = 0, R > 0\}$ would give, as in the Linear Death process, all the mass accumulating at $(0,1)$. All the other QSDs described in Theorem 4.3.2 would require initial conditions with infinite mean for them to arise as limit points.

We now prove the third point by showing that, if $\mathbf{u}$, the $(1,0)$-LCD is x-invariant, then $\mathbf{u}^A$ is also x-invariant. This follows from the system of equations $-x\mathbf{u}^T = \mathbf{u}^T Q$. More specifically, we see that for $(1,r) \in S^*$ we have, in the case where $\Phi > 0$ that

$$-xu^A(1,r) = -x\frac{u(1,r)}{\Phi} = \left(\frac{1}{\Phi}2\gamma u(2,r-1) + \delta u(1,r+1) - u(1,r)(\beta+\gamma+\delta r)\right)$$

using the fact that $u(i,r) = u^A(i,r)/\Phi$ for $(i,r) \in S^*$. This can be iterated for $i > 1$. Since the right hand side does not contain any $u(0,r)$ we can divide each term by Φ and obtain x-invariance for $\mathbf{u}^A$ conditional on $\{I > 0\}$. We have, by Theorem 4.3.4 that $\mathbf{u}^A$ is the $(1,0)$-LCD conditional on $\{I > 0\}$ and by Propositions 4.2.3 and 4.2.4 that this QSD is $(\gamma - \beta)$-invariant. As such we must have that $\mathbf{u}$ is also $(\gamma - \beta)$-invariant.

If $\Phi = 0$ then, by the second point, the QSD must have full mass on $(0,1)$ and by the x-invariance equations we must have that this is δ-invariant. Since $\Phi = 0$ if and only if $\gamma - \beta \geq \delta$, the result as stated follows. □

Remark 4.3.6. It should be noted here that we do not specifically identify a decay parameter for $\mathbf{Y}(t)$ conditional on $\{\mathbf{Y} \neq (0,0)\}$. In the Linear BDP, $\mathbf{v}$-LCDs with finite support $\mathbf{v}$ correspond to the α-invariant QSD, and the other QSDs are characterised by their invariance value. Unfortunately, we cannot make such a characterisation in this case since, in addition to the probable "high-energy" QSDs (QSDs with invariance value strictly less than the decay parameter) with mass on

S^* defined in (4.1.1), we also have the collection of "high-energy" QSDs with mass on $S \setminus S^*$ as in (4.3.3) in Theorem 4.3.2. This means that even if $\gamma - \beta < \delta$ we still have a QSD for each $x \in (0, \delta]$, contradicting the intuition conveyed in the other cases (finite or irreducible processes).

Illustration of Examples

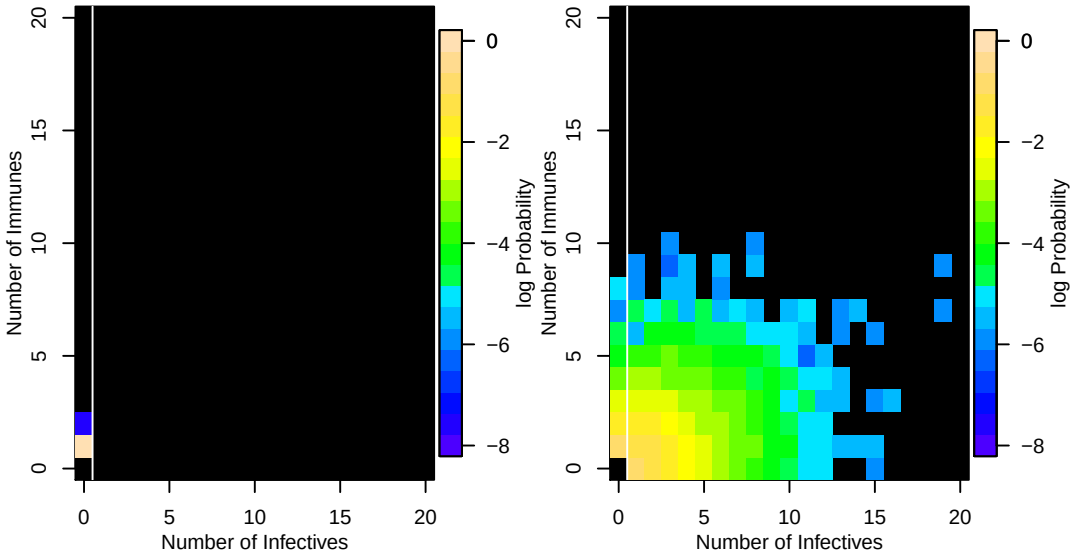
We now make use of techniques which will be developed in Chapter 5 to produce graphs of the QSDs for different values of the parameters, corresponding to the finite-support initial condition limiting conditional distributions.

Here in Figure 4.3.2 we see illustrated the $(1, 0)$-LCDs for the Transient Immunity Process conditional on $\{\mathbf{Y} \neq (0, 0)\}$ under different conditions. We can clearly see the fact that full weight ends up on $(0, 1)$ in the case where $\delta < \gamma - \beta$. This figure was produced using SMC Samplers (Section 5.1.1) with Regional Combine-Split Resampling (Section 5.5) with $M = 1000$ particles. Due to the Monte Carlo nature of the simulation techniques using a finite number of particles, we cannot obtain a perfect visualisation of the LCD which gives weight to all states, but it should be noted that, since all the states communicate in S^* they must all have strictly positive weight.

If we compare these to the QSDs conditional on $\{I(t) > 0\}$ in Figure 4.3.3 we see that even in the first case we still get weight on all states of S^* due to the communicating property. In this conditioning there is a single behaviour regime which changes continuously as the parameters change. As discussed in Remark 4.2.13, if $\delta \to 0$, then expected number of immune individuals grows to infinity, and as $\delta \to \infty$ we instead get $\mathbb{E}[R] \to 0$.

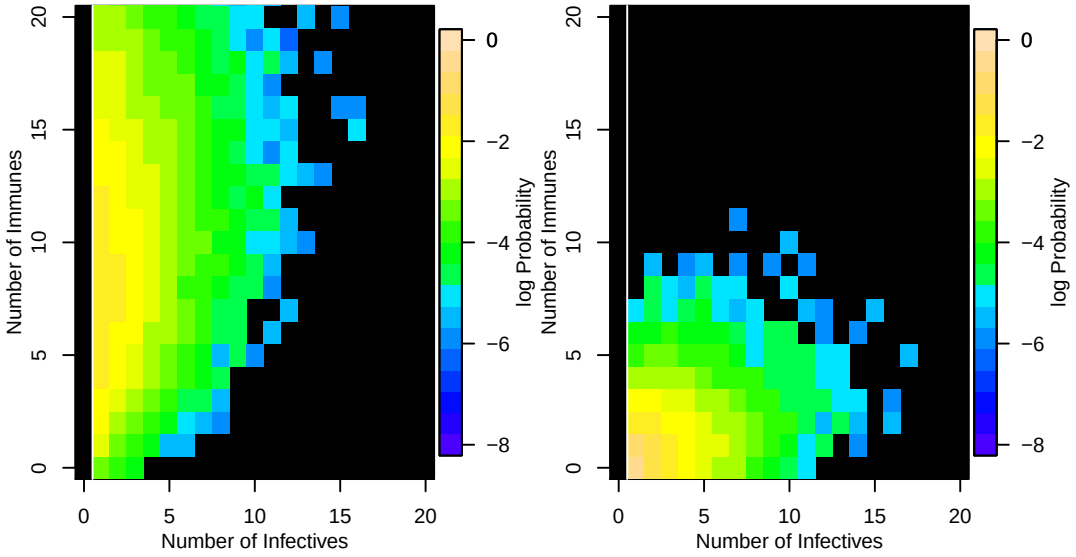
In the case of conditioning on the events $\{\mathbf{Y} \neq (0, 0)\}$, we don't have explicit expressions for the mean of the $(1, 0)$-LCD. However, we can use numerical methods to make observations. In Figure 4.3.4 we can more explicitly see the expected behaviour of the Transient Immunity Process under this conditioning with different values of β and δ; γ being fixed equal to 1. This demonstrates the fact that for $\delta < \gamma - \beta$ the expected value converges to exactly $(0, 1)$. However, as $\gamma - \beta$ decreases the process converges to a higher number of infectives, and as δ increases we see the number of expected immunes converge to lower and lower levels. As δ converges down to $\gamma - \beta$ this expected value increases. This figure was generated by numerically

solving the differential equations associated to $P_B(t)$.



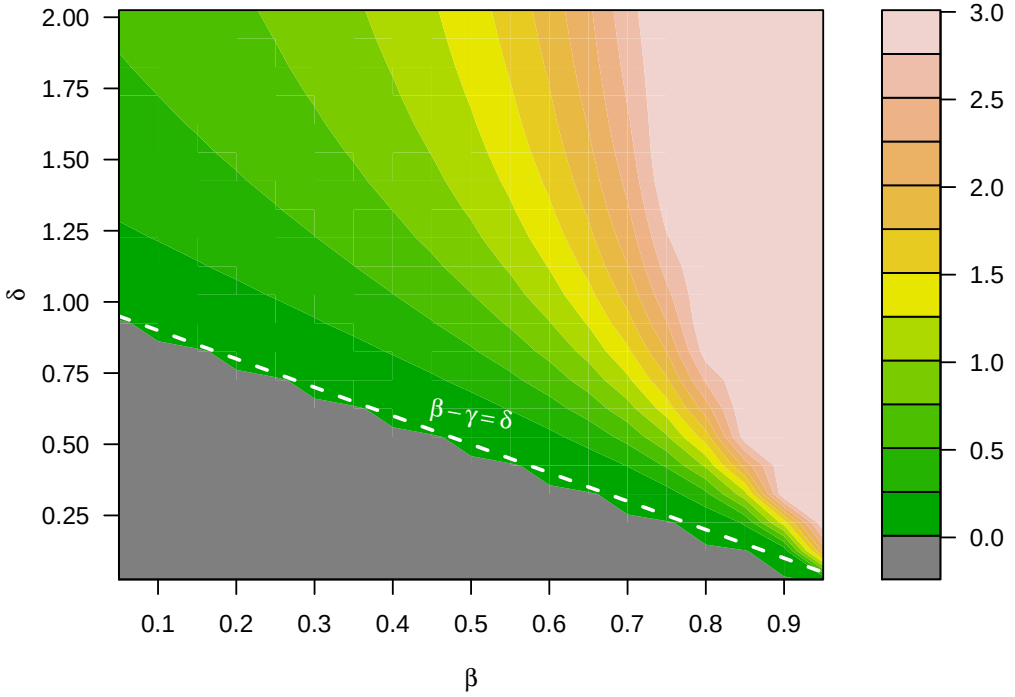
(a) $\beta = 0.5$, $\gamma = 1$, $\delta = 0.2$ (b) $\beta = 0.5$, $\gamma = 1$, $\delta = 2$

Figure 4.3.2: Visualisation of the $(1, 0)$-LCD conditional on $\{\mathbf{Y} \neq (0, 0)\}$. Simulations produced using SMC Sampler:
$M = 1000$, $T_{\max} = 20$, $T_b = 100$, $T_d = 0.25$, $T_{\text{end}} = 140$, $\lambda = 0.4$

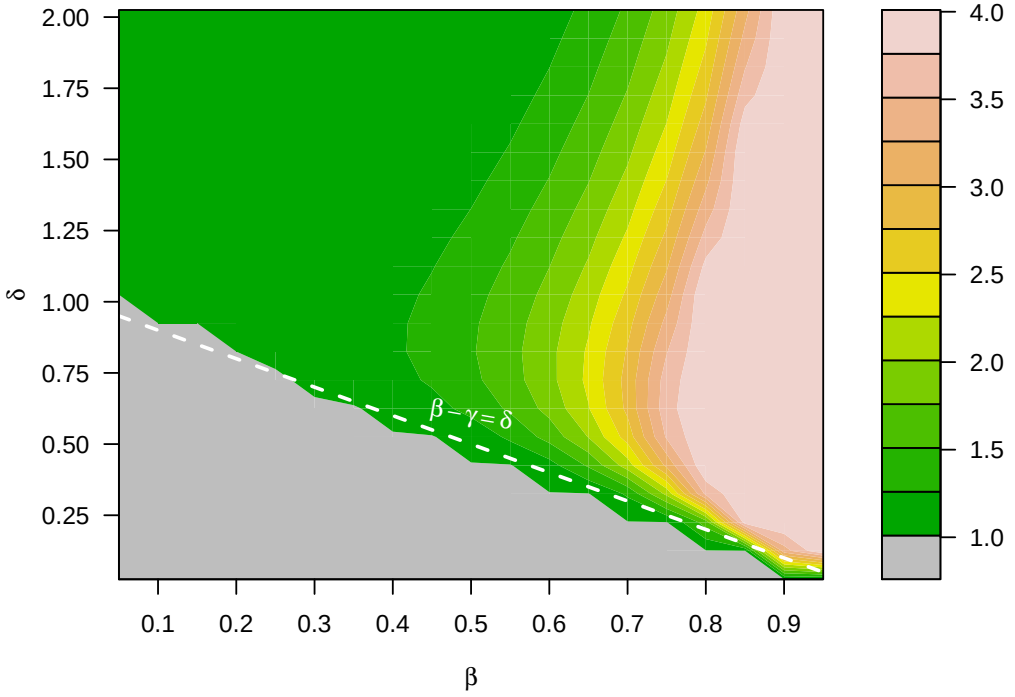


(a) $\beta = 0.5$, $\gamma = 1$, $\delta = 0.2$ (b) $\beta = 0.5$, $\gamma = 1$, $\delta = 2$

Figure 4.3.3: Visualisation of the $(1, 0)$-LCD conditional on $\{I > 0\}$. Simulations produced using SMC Sampler:
$M = 1000$, $T_{\max} = 20$, $T_b = 100$, $T_d = 0.25$, $T_{\text{end}} = 140$, $\lambda = 0.4$



(a) Number of infectives conditional on $\{\mathbf{Y} \neq (0,0)\}$



(b) Number of immune individuals conditional on $\{\mathbf{Y} \neq (0,0)\}$

Figure 4.3.4: Limiting behaviour of the Transient Immunity Process conditional on $\{\mathbf{Y} \neq (0,0)\}$ as β, δ vary with $\gamma = 1$. Obtained using numerical methods on ODEs

Chapter 5

Simulation of Quasi-Stationary Distributions

In this chapter we focus on the problem of simulating from the limiting conditional distribution of a stochastic process on a countable state space. We make use of the Sequential Monte Carlo (SMC) Sampler as defined in Chopin [2002]; Doucet et al. [2001]; Liu and Chen [1998] and discuss several novel resampling steps which make SMC sampling for QSDs and LCDs more efficient. In particular we address the difficulties that arise when sampling from processes that have a reducible state space.

Firstly, we will define the SMC Sampler, referring to concepts in Section 2.3, and how it can be used to draw from a Limiting Conditional Distribution. We focus on the class of $\mathbf{v}$-LCDs for $\mathbf{v}$ with finite support. Then we introduce the notion of refilling as a streamlined version of resampling, which reduces the computation power needed for a resampling step and could reduce variance of the resampling. The SMC sampler is then demonstrated using the multi-type Wright-Fisher model, comparing it to standard rejection sampling. Following this, we introduce the notion of Combine-Split Resampling, which preserves particle diversity in order to better explore tails of a distribution. This is demonstrated by comparing Multinomial Resampling to Combine-Split with a refilling step in simulating the 1-LCD for the linear birth-death process as defined in Chapter 3.2. Thirdly, we introduce the notion of Regional Resampling, which is our key method of tackling problems with SMC methods on reducible state spaces. We give a proof of concept using the 2-state death

process, defined in Chapter 3.3 and show convergence in Wasserstein distance. The power of Regional Resampling is then demonstrated by considering a larger death process, as well as the Transient Immunity process discussed in Chapter 4. Finally, we apply stopping time resampling schemes, similar to those mentioned by Jenkins [2012], to ensure successful applications of the above resampling algorithms, by constructing resampling steps to occur before total particle extinction, after which resampling cannot be performed, and no draw can be made without restarting the process. An application of the above techniques is then demonstrated: estimating the invariance value of the $(1, 0)$-LCD for the Transient Immunity process.

Although the above methods are developed for drawing from LCDs, they can be applied to more general situations when one might use SMC or MCMC methods to draw from distributions.

Previous Work

Some work has already been done on simulating LCDs using different methods. Groisman and Jonckheere [2013] considers a renewal process where, instead of being absorbed, particles are redrawn immediately from a given distribution μ on the transient states. They further suggests a supercritical multi-type branching process, which can be used to simulate a LCD on a finite state space through the use of a Kesten-Stigum theorem. Another paper by Blanchet et al. [2013] uses a different approach. A single particle is run to absorption, then a new particle is restarted, drawn from histogram of occupation times for the transient states.They prove that as more particles are simulated, the distribution of occupation times converges to the LCD. This is similar to the dynamical approach that Ferrari et al. [1995] uses to prove existence of QSDs in certain cases.

5.1 Sequential Monte Carlo Samplers

In this section, we take the idea of Sequential Monte Carlo defined in Chapter 2.3, and apply MCMC methodology as described by Del Moral et al. [2006] to more efficiently draw from a sequence of probability distributions. This will then be applied to the case of drawing from limiting conditional distributions.

5.1.1 Definition and Convergence

Sequential Monte Carlo Samplers provide a means to sample from a sequence of target measures $\{\pi_n : 1 \leq n \leq N\}$ over some common measurable space $(\Omega, \mathcal{F})$. Typically the target measures are only known up to some normalization constant Z_n and so it is common to work with the unnormalised measure $p_n = Z_n \pi_n$. This is done using the framework of Sequential Importance Sampling as described in Chapter 2.3, and then applying knowledge about Markov Chain Monte Carlo to construct a systematic way of drawing from one distribution given draws from the previous one.

Sampling is done using a sequence of proposal measures $\{\eta_n : 1 \leq n \leq N\}$ on Ω. Given an initial proposal measure η_1, we construct subsequent proposal measures using

$$\eta_n(x_n) = \int_\Omega \eta_{n-1}(x_{n-1}) K_n(x_{n-1}, x_n) dx_{n-1}$$

for some sequence of Markov kernels $K_n : \Omega \times \mathcal{F} \to [0, 1]$. Under naive importance sampling we would give each particle the unnormalised importance weight $w_n(x_n) = p_n(x_n)/\eta_n(x_n)$. However, such proposal distributions η_n have the possibility to become very difficult to compute pointwise as n increases, particularly if Ω is high-dimensional. To tackle this we use the following SMC sampler as described in Del Moral et al. [2006]. We define a sequence of artificial "backwards-in-time" Markov kernels $L_{n-1}(x_n, x_{n-1}) : \Omega \times \mathcal{F} \to [0, 1]$ then perform importance sampling using joint proposal distributions $\eta_n(\mathbf{x}_{1:n})$ to estimate an artificial target joint distribution $\widetilde{\pi}_n(\mathbf{x}_{1:n})$ on Ω^n defined by

$$\widetilde{\pi}_n(\mathbf{x}_{1:n}) := Z_n^{-1} \widetilde{p}_n(\mathbf{x}_{1:n})$$

where

$$\widetilde{p}_n(\mathbf{x}_{1:n}) := p_n(x_n) \prod_{k=1}^{n-1} L_k(x_{k+1}, x_k)$$

We use the same notation $\mathbf{x}_{1:n} = (x_1, \ldots, x_n)$ as in Chapter 2.3. Our final target distribution $\pi_n(x_n)$ is a marginal of our artificial target by construction, since $Z_n^{-1} p_n(x_n)$ appears explicitly in the expression. Assuming we can evaluate $\eta_1(x_1)$ and $p_1(x_1)$ to obtain unnormalised weights $w_1(x_1) = \frac{p_1(x_1)}{\eta_1(x_1)}$ then for each time-point n, we move the particles forwards in time according to $K_n(x_{n-1}, x_n)$. We then use importance sampling to approximate our artificial target $\widetilde{\pi}_n(\mathbf{x}_{1:n})$ which gives unnormalised importance weights:

$$w_n(\mathbf{x}_{1:n}) = w_{n-1}(\mathbf{x}_{1:n-1}) \widetilde{w}_n(x_{n-1}, x_n) \quad \text{for } n \geq 1$$

with incremental weights

$$\widetilde{w}_n(x_{n-1}, x_n) = \frac{p_n(x_n) L_{n-1}(x_n, x_{n-1})}{p_{n-1}(x_{n-1}) K_n(x_{n-1}, x_n)} \tag{5.1.1}$$

For M particles, this gives us an unnormalised collection of weighted particles $\{(\mathbf{X}_{0:n}^j, w_n(\mathbf{X}_{0:n}^j) : j = 1, \ldots, M\}$. As in sequential importance sampling, we normalise the importance weights to get

$$W_n^j = \frac{w_n(\mathbf{X}_{0:n}^j)}{\sum_{k=1}^M w_n(\mathbf{X}_{0:n}^k)}$$

These weights can then be used to generate samples from the marginal distribution of interest $\pi_n(x_n)$.

The SMC sampler described above, and summarised in Algorithm 5.1.1, can be accompanied by a resampling scheme such as those described in Section 2.3.3.

5.1.2 Simulating LCDs using SMC samplers

To implement the SMC sampler in order to simulate from a **v**-LCD, we would ideally take the **v**-LCD itself to be the target distribution. However, in general we will not be able to compute even the unnormalised density of the LCD, and so we cannot compute importance weights within an SMC sampler scheme. Instead, we use the

Algorithm 5.1.1 SMC Sampler

Require: $M \geq 1$, $N \geq 1$, η_1 probability measure, transition kernels $K_n(x, y)$, $L_n(x, y)$ for $1 \leq n \leq N$.

1: Draw $X_1^j \sim \eta_1$ for $j = 1, \dots M$.

2: For $j = 1, \dots, M$ compute $w_1(X_1^j) = \frac{p_1(X_1^j)}{\eta_1(X_1^j)}$ and normalise to give $W_1^j := w_1(X_1^j) / \sum_{j=1}^M w_j(X_1^j)$.

3: **for** $n = 2$ **to** N **do**

4: For $j = 1, \dots, M$ draw $X_n^j \sim K_n(X_{n-1}^j, \cdot)$.

5: For $j = 1, \dots, M$ compute $\widetilde{w}_n(X_{n-1}^j, X_n^j)$ from (5.1.1) and normalise to get W_n^j.

6: **end for**

7: **return** $\pi_n^M(x) = \sum_{j=1}^M W_n^j \delta_{X_{1:n}^j}(x)$ for $n = 1, \dots, N$.

time marginal of the process in question conditional on non-absorption given by

$$\pi_T(\cdot) = \mathbb{P}_{\mathbf{v}}[X(T) \in \cdot | X(T) \in S]$$

which converges to the true $\mathbf{v}$-LCD as T gets large. It is the time marginal π_T which we will attempt to simulate using an SMC sampler.

Define an increasing sequence of time points $\{t_n : n = 1, \dots, N\}$ with $t_1 = 0$ and $t_N = T$, and set the initial proposal distribution η_1 to be $\mathbf{v}$. When simulating analytically intractable LCDs we cannot even work with an unnormalised target $p_n(x)$, and so we might as well set $\pi_n(x) = p_n(x)$ (i.e. $Z_n = 1$). We construct the sequence of proposal distributions $\eta_n(\cdot) = \mathbb{P}_{\eta_1}[X(t_n) \in \cdot]$ using the Markov transition kernel

$$K_n(x_{n-1}, x_n) = \mathbb{P}[X(t_n) = x_n | X(t_{n-1}) = x_{n-1}]$$

and the backward kernel

$$L_{n-1}(x_n, x_{n-1}) = \frac{\mathbb{P}[X(t_{n-1}) = x_{n-1}]}{\mathbb{P}[X(t_n) = x_n]} \mathbb{P}[X(t_n) = x_n | X(t_{n-1}) = x_{n-1}]$$

This precisely matches the optimal choice of backward kernel given by Del Moral et al. [2006], which minimizes the variance of the unnormalised importance weights $w(\mathbf{x}_{1:n})$,

$$L_{n-1}^{\text{opt}}(x_n, x_{n-1}) = \frac{\eta_{n-1}(x_{n-1}) K_n(x_{n-1}, x_n)}{\eta_n(x_n)}$$

Proposition 5.1.1. *If* $\mathbf{v}$ *has full mass on* S*, the particle weights at each timepoint of the SMC sampler used to simulate the* $\mathbf{v}$*-LCD are equally distributed amongst non-absorbed particles and zero for absorbed particles.*

Proof. Since $\eta_1 = \mathbf{v}$, for $n = 1$ we have that

$$w_1(x_1) = \frac{\pi_1(x_1)}{\eta_1(x_1)} = \frac{\mathbb{P}_{\mathbf{v}}[X(t_1) = x | X(t_1) \in S]}{\eta_1(x)} = 1$$

and so the particles begin with equal weights.

For $n \geq 2$, if we substitute the expressions for L_n, K_n, η_n and π_n into the incremental weight $\widetilde{w}_n$ (suppressing the conditioning on $X(t_1) \sim \mathbf{v}$ for brevity) we see that

$$\begin{aligned}
\widetilde{w}_n(x_{n-1}, x_n) &= \frac{\pi_n(x_n) L_{n-1}(x_n, x_{n-1})}{\pi_{n-1}(x_{n-1}) K_n(x_{n-1}, x_n)} \\
&= \frac{\pi_n(x_n)}{\pi_{n-1}(x_{n-1})} \frac{\mathbb{P}[X(t_{n-1}) = x_{n-1}]}{\mathbb{P}[X(t_n) = x_n]} \frac{\mathbb{P}[X(t_n) = x_n | X(t_{n-1}) = x_{n-1}]}{\mathbb{P}[X(t_n) = x_n | X(t_{n-1}) = x_{n-1}]} \\
&= \frac{\mathbb{P}[X(t_n) = x_n | X(t_n) \in S]}{\mathbb{P}[X(t_n) = x_n]} \frac{\mathbb{P}[X(t_{n-1}) = x_{n-1}]}{\mathbb{P}[X(t_{n-1}) = x_{n-1} | X(t_{n-1}) \in S]}.
\end{aligned}$$

Substituting this incremental weight into the full weight (5.1.1) gives a telescoping product which reduces to

$$w_n(x_n) = \frac{\mathbb{P}[X(t_n) = x_n | X(t_n) \in S]}{\mathbb{P}[X(t_n) = x_n]} w_1(x_1).$$

For $x_n = 0$ we immediately get $w_n(0) = 0$ whilst for $x_n \in S$, by the definition of conditional probability,

$$w_n(x_n) = \frac{1}{\mathbb{P}[X(t_n) \in S]},$$

which is independent of the particle location x_n and so the weight is equally distributed amongst the non-absorbed particles. □

Proposition 5.1.1 shows that drawing the $\mathbf{v}$-LCD via SMC sampling works in a similar way to rejection sampling, where only the non-absorbed particles are considered. The existence of a LCD requires the certain absorption of each particle in a finite amount of time and so there remains a problem balancing the approximation of the LCD (improved by increasing T) and particle depletion (worsened by increasing T). However setting our algorithm within an SMC framework allows us to draw on

existing tools to prevent particle depletion, such as particle resampling.

We should also note here that one can incorporate resampling schemes into the SMC sampler as in standard Sequential Monte Carlo. At any of the timepoints considered, we can follow the simulation step by a resampling step. In what follows, we will for the time being use a deterministic timepoint sequence where we resample at $\{T_{\text{step}}, 2T_{\text{step}}, 3T_{\text{step}}, \dots\}$ for $T_{\text{step}} > 0$, a pre-chosen parameter. This needs to be chosen such that absorption is unlikely to happen between any two timepoints. We will return to this in Section 5.5, and discuss methods to mitigate this, but for now, assume all choices of T_{step} are made so as to be as large as possible whilst having a low probability of absorption of all particles between two resampling steps. In the case where we do not implement resampling, one can consider each sampling step to behave like a rejection sampling step, where we reject all such draws which give absorbed particles. In the rest of this chapter, we will refer to the SMC sampler without resampling by *SMC rejection sampling*.

5.2 Particle Refilling

We have seen that in order to make use of SMC samplers to draw from **v**-LCDs with finite support **v**, we need only simulate forward the unconditioned process and give uniform non-zero weight to the non-absorbed particles. The typical resampling schemes are used to mitigate particle weight degeneracy, but under this method, weight degeneracy usually occurs only due to a lack of non-absorbed particles.

One should at this point note that in the finite or countable state space, one should mark the distinction between *location weight* (the sum of particle weight at a given location) and *particle weight*. In the continuous state space, it is generally the case that, with certain probability, all particles will be at different locations for all times $t > 0$. As a result, particle and location weight coincide in this setting. However, if the number of particles M is greater than the number of states, then at any timepoint one must have at least one state at which inhabit multiple particles. Even in the countable case, it is possible that multiple particles can be inhabiting a given location, particularly in cases where the proposal distributions are concentrated in a finite region within the state space.

Resampling schemes such as those described in Section 2.3.3 will often throw away particles with low importance weight and multiply those with high weight. As mentioned before, this increases weight variance. To improve the resampling step to a more computationally efficient one which also suffers from less Monte Carlo variance, we introduce a new technique which we will refer to as *Particle Refilling*. Consider a simulation which runs M particles. If at a resampling timepoint t there are $M_S(t)$ non-absorbed particles, then particle refilling leaves these particles alone, and draws $M_0(t) = M - M_S(t)$ particles from the non-absorbed, and discards the absorbed ones. Such resampling draws can be made according to any resampling method, such as those previously outlined: under the notation of Section 2.3.3 we are setting $M' = M_0(t)$, and taking the new set of particles going forward to be the union of the previously non-absorbed particles and the selection of newly resampled ones.

This refilling scheme, compared to those previously mentioned, has the benefit of never throwing away particles which are still not absorbed; such particles run for sufficiently long will be good approximations of draws from the LCD in question. We also have the added benefit of assured particle diversity since any location with particles inhabiting it before the resampling step will still have them afterwards.

This will allow potentially better exploration of the tails of a distribution or the movement of particles between distant modes of a multi-modal distribution. Furthermore, it involves the resampling of fewer particles, which reduces computation overheads. This method is formally defined in Algorithm 5.2.1.

Algorithm 5.2.1 Rejection Refilling

Require: $M \geq 1$, $v \in \{1, \ldots, N\}^M$, $T > 0$, $\{t_1 \leq \cdots \leq t_K\} \subset (0, T)$
1: Set $X^{(j)}(0) = v_j$ for $j = 1, \ldots M$.
2: Run each particle until time t_1 obtaining $X^{(j)}(t_1)$ for $j = 1, \ldots, M$.
3: Set $\mathcal{U}(t_1) \leftarrow \{j : X^{(j)}(t_1) \neq 0\}$.
4: **for** $k = 2$ **to** K **do**
5: **if** $X^{(j)}(t_{k-1}) = 0$ **then**
6: Draw x_j from $\mathcal{U}(t_{k-1})$ uniformly with replacement for $j = 1, \ldots, M$.
7: Set $X^{(j)}(t_{k-1}^+) \leftarrow x_j$.
8: **else**
9: Set $X^{(j)}(t_{k-1}^+) \leftarrow X^{(j)}(t_{k-1})$
10: **end if**
11: Run each particle until time t_k obtaining $X^{(j)}(t)$.
12: Set $\mathcal{U}(t_k) \leftarrow \{j : X^{(j)}(t_k) \neq 0\}$.
13: **end for**
14: **return** $\{X^{(j)}(T) : j \in \mathcal{U}\}$

Recalling Definition 2.3.1, we show that refilling preserves proper weighting.

Proposition 5.2.1. *Given a properly weighted sample $\{(X_j, w_j) : 1 \leq j \leq M\}$ with respect to some measure π and a resampling method that produces properly weighted samples then particle refilling, where only the zero weight particles are resampled from the non-zero weight particles using the given resampling method, produces properly weighted samples.*

Proof. Let (X'_j, w'_j) denote the particle location and weight after particle refilling. Then conditioning on $X_j = 0$ (or equivalently $w_j = 0$) yields $\mathbb{E}[h(X'_j)w'_j] = \mathbb{E}[h(X_K)w_K]\mathbb{P}[X_j = 0] + \mathbb{E}[h(X_j)w_j]\mathbb{P}[X_j \neq 0] = c\mathbb{E}_\pi[h(X)]$, where K is the index of the resampled particle randomly chosen from the non-zero weight particles using the given resampling method. □

Taking multiple samples

When simulating LCDs using SMC samplers, we can take advantage of the fact that after a suitable burn-in period, every target distribution π_n is an approximation of

the true LCD. As a result, we borrow ideas from MCMC and draw samples from many timepoints after the SMC sampler has reached stationarity and not just the final timepoint. In what follows, we adopt a burn-in period $T_b \geq 0$, during which the samples are discarded. If one were to start from the LCD of interest, one could theoretically take $T_b = 0$, but in practise, this burn-in time can only be chosen meaningfully by observing a trace of the simulation and adjusting.

One issue that arises in the case of taking multiple samples from the same simulation is one of independence. Typically, the standard Monte Carlo integration results rely on independence of all the samples. By taking multiple samples from the same set of particles, one must observe that the particles taken later in time through the simulation may be dependent on the state of the process at the time of the previous observation. Even in the Markovian case, each observation taken is generally dependent on the previous one. In many cases, one can make the assumption that given a long enough time between observations, such observations have negligible autocorrelation. We implement this assumption by stipulating thinning of the observations in the form of a delay $T_d > 0$ between sampling times. Again, such a delay parameter cannot be determined analytically; the optimal delay being $T_d = \infty$. in practice, one observes the autocorrelation of the process and chooses T_d appropriately, and then rerunning the simulation.

5.2.1 The Two-type Wright-Fisher Model

In order to demonstrate the basic SMC sampler, we apply it to simulate a LCD of the discrete-time Wright-Fisher process. Population genetics forms an important application area for QSDs and LCDs, where they represent the distribution of existing genetic types within a population.

Some genetic variations occur at just one base on the genome and are known as single nucleotide polymorphisms (SNPs). At each SNP there may be one of four nucleotides present – Adenine, Thymine, Cytosine or Guanine. This we could instead consider as a population of SNPs each of which has a genetic type $\{A, T, C, G\}$, which change over generations of replication. More generally, consider a population of D haploid individuals (individuals for which each offspring has only one parent), each with an allelic type from a set of $K > 1$ types denoted $\{1, \ldots, K\}$. At each timepoint $n \in \mathbb{N}$, we generate from the current population a new set of offspring, which will form the population of the next generation. If the allelic types confer no advantage, the

process is referred to as *neutral*, and each offspring independently chooses a parent uniformly at random and adopts the allelic type of the parent. More generally, each allelic type k is assigned a *selection coefficient* $s_k > -1$ where each offspring selects a particular parent with type k with probability proportional to $s_k + 1$. This *Wright-Fisher with Selection* process, as described in Tavaré [1984] evolves over a finite state space of D^K states.

One should note that for finite D and K the process will, in finite time, reach a point at which there will only be a single allelic type in the population, since once one generation has no individuals of a given type, that type cannot appear in any future generations. It should be noted that one could also incorporate mutation mechanisms into the Wright-Fisher process whereby individuals can mutate from the parental type to a different one according to a mutation transition matrix. In this case, if the mutation matrix is irreducible, then the process would be recurrent, since it would always be possible to return from a single-type population to one with multiple types.

We focus on the 2-type Wright-Fisher process and consider the quasi-stationary distribution conditioned on the event that there are still 2 allelic types in the population. Since this evolves on a finite irreducible state space, we get from Proposition 2.1.18 that there is a unique QSD conditional on the event that both types remain in the population, and this is also the unique LCD, starting from any distribution containing both types. In particular we consider the $\mathbf{v}$-LCD for $\mathbf{v}$ which gives full weight to the state with $D/2$ individuals of each of the two types. We can find the true $\mathbf{v}$-LCD as a left eigenvector of the transition matrix, however for large population sizes, this is numerically demanding. Here we use a small enough population to generate the true LCD for illustrative purposes.

We compare the SMC sampler scheme with multinomial refilling to the basic SMC rejection sampling algorithm. Here we set resampling to occur every 5 timepoints, so $T_{\text{step}} = 5$ as defined previously. In Figure 5.2.1, the main problem with rejection sampling becomes evident: the number of accepted particles decreases rapidly through time. In comparison, the SMC sampler replenishes the particles at each resampling step. Figure 5.2.2 shows the estimated LCD of the number of individuals of type 1 from the two methods, compared against the true QSD computed using `eigen` in R. It should be noticed that due to the small number of non-absorbed particles, the SMC Rejection Sampler performs poorly giving weight to only three states in the process. Conversely, we see that even with a small number of particles,

with Multinomial Refilling we obtain a reasonable estimate for the LCD. One could obtain a better estimate with more samples taken or more particles evolving but for direct comparison we only show this case.

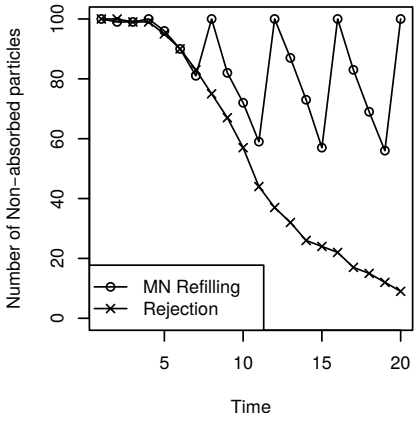
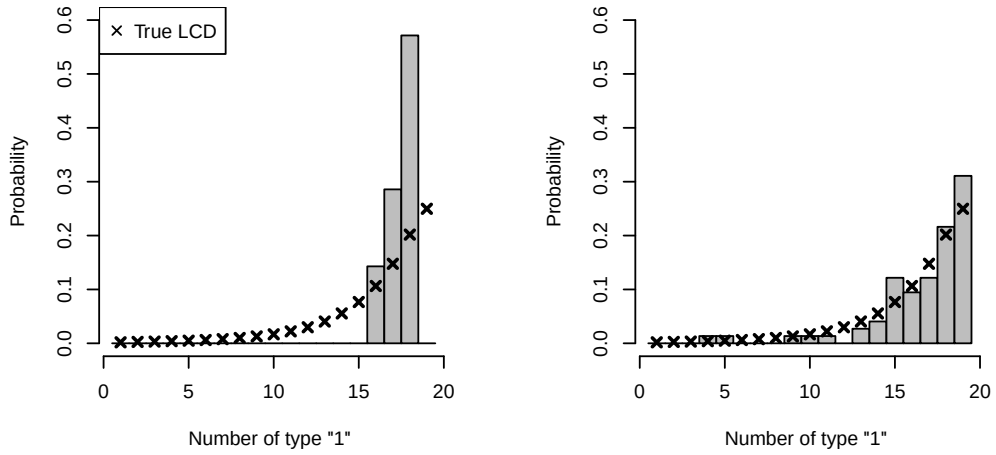


Figure 5.2.1: Comparison of the number of non-absorbed particles for the SMC Multinomial Refilling sampler and the SMC rejection sampler; $D = 20$, $\mathbf{s} = (0, 0.1)$, $M = 100$.



(a) SMC Rejection sampler (b) Multinomial Refilling SMC Sampler

Figure 5.2.2: LCD for a Wright-Fisher process with selection, simulated using a rejection sampler and an SMC sampler;
$D = 20$, $M = 100$, Simulation time $T_{\text{end}} = 20$, $T_{\text{step}} = 5$, $\mathbf{s} = (0, 0.1)$.

5.3 Combine-Split Resampling

Traditional SMC resampling schemes are designed for particles evolving in continuous state spaces where no single point has strictly positive mass. As previously mentioned, in discrete state spaces it is likely that several particles share the same location. We propose the *Combine-Split* resampling method which redistributes particles within the space without moving any of the weight, with the aim of improving the effective sample size. One scenario in which this is particularly useful is when the absorbing state can only be accessed from one of the transient states, which is typically the mode of the LCD. As a result, one tends to find either far too many particles in this state, which leaves few to explore the tails, or one very heavy particle as too many other particles have just been absorbed and hence have no weight at all. Combine-Split attempts to rectify this using the following method.

5.3.1 Idea and algorithm

Combine split resampling comprises three steps, as seen in Algorithm 5.3.1. First, at each state $s \in S$ we combine together all the particles which are at that location s into a single particle and give it the combined weight of all the particles that were sitting there. Next, all non-assigned particles (those lost in the combining and the absorbed particles) are distributed amongst the locations with non-zero weight according to some chosen distribution, and are assigned temporary weight zero. Finally all of the particles now residing at a given location are given weight equal to the total weight at that location divided by the number of particles there.

Algorithm 5.3.1 Combine-split resampling

Require: $M \geq 1$, $\{(X_i, W_i) : 1 \leq i \leq M\}$ normalised weighted particles.

1: For each $s \in S$ let $a_s = \sum_{i=1}^{M} W_i \mathbb{1}_{\{X_i = s\}}$ and $S^* = \{s \in S : a_s > 0\}$.
For each $s \in S^*$ let one particle $X_i = s$ have weight a_s. Give all other particles zero weight.

2: For each zero-weight particle, draw from some specified distribution over S^*.

3: For each $s \in S^*$ let $N_s = |\{X_i = s\}|$.
For $i = 1, \ldots, M$ set $W_i = a_s / N_s$ where $s = X_i$.

4: **return** $\{(X_i, W_i) : 1 \leq i \leq M\}$ normalised weighted particles

The weight at each state remains constant during combine-split resampling, the particles are simply redistributed amongst those states. All of the other resampling mechanisms described in Section 2.3.3 can lose particle locations and therefore re-

Particle name	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
Initial Location	a	a	a	b	b	c	0	0
Initial Weight	1	1	2	1	4	2	0	0
Combined Location	a	0	0	b	0	c	0	0
Combined Weight	4	0	0	5	0	2	0	0
Reallocated Location	a	a	b	b	b	c	c	c
Reallocated Weight	4	0	0	5	0	2	0	0
Split Location	a	a	b	b	b	c	c	c
Split Weight	2	2	5/3	5/3	5/3	2/3	2/3	2/3

Table 5.3.1: Particles' locations and weights during combine-split resampling step

duce the diversity of the particles. We prove proper weighting and then illustrate the method.

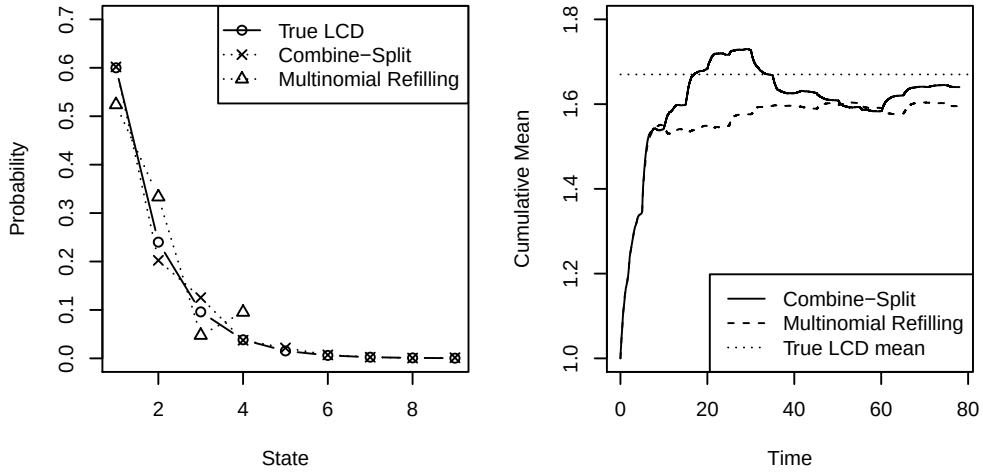
Proposition 5.3.1. *Given a properly weighted sample* $\{(X_j, w_j) : 1 \leq j \leq M\}$ *with respect to some measure* π*, combine-split resampling produces a properly weighted sample.*

Proof. Let $\{(X'_j, w'_j)\}$ denote the locations and weights after combine-split resampling. Since combine-split does not change the total weight at any location, we must have $\mathbb{E}[h(X'_j)w'_j] = \mathbb{E}[h(X_j)w_j] = c\mathbb{E}_\pi[h(X)]$, and so the new sample is properly weighted. □

Example 5.3.2. Suppose that there are eight particles with locations and weights as given in rows 2 and 3 respectively of Table 5.3.1, of which two have been absorbed into state zero. The combine step (rows 4 and 5) moves the weight at locations a, b and c to a single particle at each location, leaving three extra particles to reallocate (five in total). Suppose that we reallocate these 5 particles by drawing uniformly at random from the three locations, giving (a, b, b, c, c), as seen in rows 6 and 7. Finally, the *Split* step equalises the weight at each location, as shown in rows 8 and 9.

5.3.2 Application: Linear birth-death process

In this example we compare the true LCD for the Linear BDP as explored in Chapter 3 with the simulated 1-LCD produced using an SMC sampler with multinomial refilling and with combine split resampling. In the combine-split resampling step,



(a) True and estimated LCDs (b) Cumulative mean of particles

Figure 5.3.1: The LCD for a linear birth-death process:
$\beta = 0.4$, $\gamma = 1$, $M = 100$, Simulation time $T_{\text{end}} = 80$, $T_{\text{step}} = 4$, $T_b = 40$, $T_d = 2$.

the zero-weight particles were reallocated to locations drawn uniformly at random from the existing locations, which sends more particles to the tail than reallocating proportionally to the weights. Figure 5.3.1a shows the true and estimated LCDs. It is clear that the SMC with combine-split resampling reaches further into the tail than SMC with multinomial refilling, and hence matches the true LCD more closely. In Figure 5.3.1b, we see the cumulative mean of the particles observed at each resampling step. The cumulative mean under combine-split is consistently higher than under multinomial refilling and closer to the true mean. This suggests faster convergence to the true LCD and a reduction in the finite sample bias inherent in particle approximation methods.

Remark 5.3.3. The Combine-Split Algorithm (Algorithm 5.3.1) is very flexible because there still is the choice of how one might reallocate the zero-weight particles. Two obvious options are proportional and uniform weighting. Allocating particles to locations with probability proportional to the location weight will better split the particles up to reduce the loss of weight due to absorption, and will increase the effective sample size due to more evenly splitting the weight between particles. On the other hand, allocation uniformly at random allows more particles to be given to the tails on average, which will allow a better exploration of the tails at the cost of larger weight particles in locations close to local or global maxima of the LCD.

5.4 Regional Resampling

One difficulty with SMC sampling on a reducible state space is that once all the particles have left a transient communicating class there is no mechanism for particles to return there. Since the support of the initial distribution determines which of the LCDs is being estimated, this changes the LCD that the SMC sampler is converging towards so that it is no longer the required target distribution. To address this weakness, we propose regional resampling in which the state space is partitioned into suitable regions and the number of particles available to explore each region is stipulated in advance. At each resampling timepoint, particles are removed from regions with too many particles and reallocated to regions with too few. Although we describe regional resampling in the context of LCDs, we anticipate that it will have wider applications in more general SMC schemes.

5.4.1 Notation and Algorithm

Let $X = (X(t))_{t\geq 0}$ be an absorbing Markov process on a state space $S \cup \{0\}$, with absorbing state 0. We partition the transient states into L further regions; $S = \bigcup_{l=1}^{L} S_l$. We specify $\mathbf{N} = (N_1, N_2, \ldots, N_L) > 0$ to be the desired number of particles in each region (or *populations*), such that $\sum_{l=1}^{L} N_l = M$, where M is the number of particles used in the simulation. We again make use of a deterministic resampling timepoint sequence $\{T_{\text{step}}, 2T_{\text{step}}, \ldots\}$ as before.

Each resampling step begins with a set of normalised weighted particles $\{(X_i, W_i) : i = 1, \ldots, M\}$ Any absorbed particles are defined to have weight zero. Let $M_l(t)$ be the number of particles in region l at resampling time t and define $W(l)$ to be the total weight in region l; $W(l) = \sum_{i=1}^{M} W_i \mathbb{1}_{\{X_i \in S_l\}}$.

In regional resampling, we resample particles in each region separately. In region l, we draw N_l particles from the $M_l(t)$ existing particles in that region using any resampling algorithm (such as Combine-Split, particle refilling or an existing algorithm as described in Section 2.3.3 with $M' = N_l$). This region-specific resampling will be referred to as *reallocation* to distinguish it from the resampling taking place on the whole transient state space S. The weights of these new particles are then renormalised so that the total weight in region l remains equal to $W(l)$. For example, after multinomial resampling within region S_l, the renormalised weights of each particle would be $W(l)/N_l$.

In applications to sampling from LCDs we expect that $N_l \geq M_l(t)$ for all l in most cases, because some of the particles will have been absorbed since the last resampling step. However, it is possible that in some regions $N_l < M_l(t)$ and so if we are applying combine-split or particle refilling it may not be possible to keep all of the locations. It is therefore necessary to apply an alternative resampling step such as multinomial resampling in these instances. The regional resampling algorithm is described in Algorithm 5.4.1.

Algorithm 5.4.1 Regional Resampling

Require: A partition $S = \bigcup_{l=1}^{L} S_l$; $\{(X_i, W_i) : 1 \leq i \leq M\}$ normalised weighted particles.

1: Set $W(l) = \sum_{i=1}^{M} W_i \mathbb{1}_{\{X_i \in S_l\}}$ be the weight in region l.
2: For $l = 1, \ldots, L$, resample N_l particles from region S_l then renormalise the weights within the region to have total $W(l)$.
3: **return** $\{(X_i, W_i) : 1 \leq i \leq M\}$ normalised weighted particles.

It should be noted that one does not need to choose S_l to be a single transient communicating class. This method can be extended to any choice of $\{S_l\}_{l=1}^{L}$ as long as they form a complete partition of the transient states.

Proposition 5.4.1. *Given a properly weighted sample $\{(X_j, w_j) : 1 \leq j \leq M\}$ with repsect to some measure π then regional resampling, using a resampling method that produces properly weighted samples within each region, produces properly weighted samples.*

Proof. Let (X'_j, w'_j) be the location and weight of particle j after resampling. Then since resampled particles are properly weighted within each region, $\mathbb{E}[h(X'_j)w'_j] = \sum_{l=1}^{L} \mathbb{E}[h(X'_j)w'_j \mathbb{1}_{\{X'_j \in S_l\}}] = \sum_{l=1}^{L} c_l \mathbb{E}_\pi[h(X)\mathbb{1}_{\{X \in S_l\}}] = c\mathbb{E}_\pi[h(X)]$, where $c = \sum_{l=1}^{L} c_l$ for all j. Hence regional resampling produces properly weighted particles. □

5.4.2 Proof of convergence for a simple example

In general it is extremely challenging to prove convergence of such SMC sampler methods with resampling, except in very specific circumstances. We present here a simplified model to illustrate that the algorithm does indeed converge in some situations.

Consider the pure death process on $\{0, 1, 2\}$. Transitions from state 2 to state 1

occur with rate δ and transitions from state 1 to 0 (absorption) occurs with rate 1. When $0 < \delta < 1$, the 2-LCD is given by $(\delta, 1-\delta)$; when $\delta \geq 1$ the 2-LCD is simply $(1, 0)$. We wish to prove that the SMC sampler with regional resampling converges. Choose two regions given by $S_1 = \{1\}$, $S_2 = \{2\}$. We stipulate N_1, N_2 with $N_1 + N_2 = M$ to be the desired populations of each region. In what follows, we let $w_l(t)$ to be the unnormalised weights on the particles in state l at time t. We want to show that the normalized weights $(W_1(t), W_2(t))$ do indeed converge to some distribution for any choice of $(W_1(0), W_2(0))$, if we take the limit $t \to \infty$ at the sequence of resampling times $(t_n)_{n\geq 1}$.

We look at the simplest case where combine-split resampling happens within each region after every event in the simulation. In this case, we can see the unnormalised weights as moving at points of a Poisson process with rate $N_1 + \delta N_2$, with the jump chain $(w_1, w_2)(n) = (w_1(t_n), w_2(t_n))$ moving like

$$(w_1, w_2)(n+1) = \begin{cases} \left(w_1(n)\frac{N_1-1}{N_1}, w_2(n)\right) & \text{w.p.} \quad \frac{N_1}{N_1+\delta N_2}, \\ \left(w_1(n) + \frac{w_2(n)}{N_2}, w_2(n)\frac{N_2-1}{N_2}\right) & \text{w.p.} \quad \frac{\delta N_2}{N_1+\delta N_2}. \end{cases}$$

Each jump immediately triggers a combine-split resampling within each region. In this model the regions consist of just one state, and so the combine-split resampling simply spreads the weight uniformly amongst the particles in that region.

If we define $X(n) := W_1(t_n)$, this Markov chain evolves according to:

$$X(n+1) = \begin{cases} \frac{X(n)(N_1-1)}{N_1-X(n)} & \text{w.p.} \quad \frac{N_1}{N_1+\delta N_2} \\ X(n) + \frac{1-X(n)}{N_2} & \text{w.p.} \quad \frac{\delta N_2}{N_1+\delta N_2} \end{cases} \tag{5.4.1}$$

Since we know that $W_2(t_n) = 1 - X(n)$, if we can show convergence in Wasserstein distance of $X(n)$ then we have convergence in Wasserstein distance of both weights.

Definition 5.4.2. Let E be a Polish space, and let $d : E \times E \to [0, 1]$ be a distance-like function (symmetric, lower-semi-continuous and $d(x, y) = 0 \Leftrightarrow x = y$). Then the *Wasserstein-d distance* between two probability measures μ, ν on E is given by

$$\mathcal{W}_d(\mu, \nu) = \inf_{\pi} \int_{E\times E} d(u, v)\pi(du, dv)$$

where π runs over all probability measures on $E \times E$ which have marginals μ, ν.

Theorem 5.4.3. *The distribution of $X(n)$ on $(0, 1)$ defined in (5.4.1) converges in*

Wasserstein-d distance with

$$d(x,y) := \min\left\{\left|\frac{1}{1-x} - \frac{1}{1-y}\right|, 1\right\}$$

to some stationary distribution π, for all $\delta \in (0,1)$, $N_1 \geq 2$, $N_2 \geq \max(5, \frac{1}{1-\delta})$.

We prove this be applying the following theorem taken from Hairer et al. [2014].

Theorem 5.4.4. *Let P be a Markov kernel over a Polish space E and assume that:*

1. *P has a Lyapunov function $V : E \to \mathbb{R}$ such that there exists $\lambda \in [0,1)$ and $K > 0$ such that*

$$PV(x) := \int_E V(u)P(x,du) < \lambda V(x) + K$$

   for all $x \in E$.

2. *P is d-contracting for some distance-like function $d : E \times E \to [0,1]$ (d is symmetric, lower-semi-continuous and $d(x,y) = 0 \Leftrightarrow x = y$), so that there exists $c \in (0,1)$ such that for every $x, y \in E$ where $d(x,y) < 1$ we have*

$$\mathcal{W}_d(P(x,\cdot), P(y,\cdot)) < c\, d(x,y).$$

3. *The set $S = \{x : V(x) < 4K\}$ is d-small, so that there exists $s \in (0,1)$ such that for all $x, y \in S$*

$$\mathcal{W}_d(P(x,\cdot), P(y,\cdot)) \leq s.$$

Then there exists $n \in \mathbb{N}$ such that for any two probability measures μ, ν on E we have

$$\mathcal{W}_{\tilde{d}}(\mu P^n, \nu P^n) \leq \mathcal{W}_{\tilde{d}}(\mu, \nu)$$

where $\tilde{d}(x,y) = (d(x,y)(1+V(x)+V(y)))^{1/2}$, and n is increasing in λ, K, c, s. Hence there is at most one invariant measure.

Moreover, if the following hold:

4. *There exists a complete metric d_0 on E such that $d_0 \leq \sqrt{d}$*

5. *P is Feller on E, which holds precisely when for any continuous function f on E, $\int_E f(y)P(x,dy)$ is continuous for every $x \in E$,*

then there exists a unique invariant measure μ for P.

In order to do this we will make use of the following lemmas.

Lemma 5.4.5. *The function $V(x) = x(1-x)^{-1}$ is Lyapunov for P as above with*

$$\lambda = 1 - \frac{N_2(1-\delta) - 1}{(N_2 - 1)(N_1 + \delta N_2)}$$
$$K = \frac{\delta N_2}{(N_2 - 1)(N_1 + \delta N_2)}$$

Lemma 5.4.6. *P is d-contracting for the distance-like function*

$$d(x, y) = \min\left\{1, \left|(1-x)^{-1} - (1-y)^{-1}\right|\right\}$$

Lemma 5.4.7. *The set $S = \{x : V(x) < 4K\}$ is d-small with $V(x)$ and K as defined in Lemma 5.4.5, and $d(x, y)$ as in Lemma 5.4.6.*

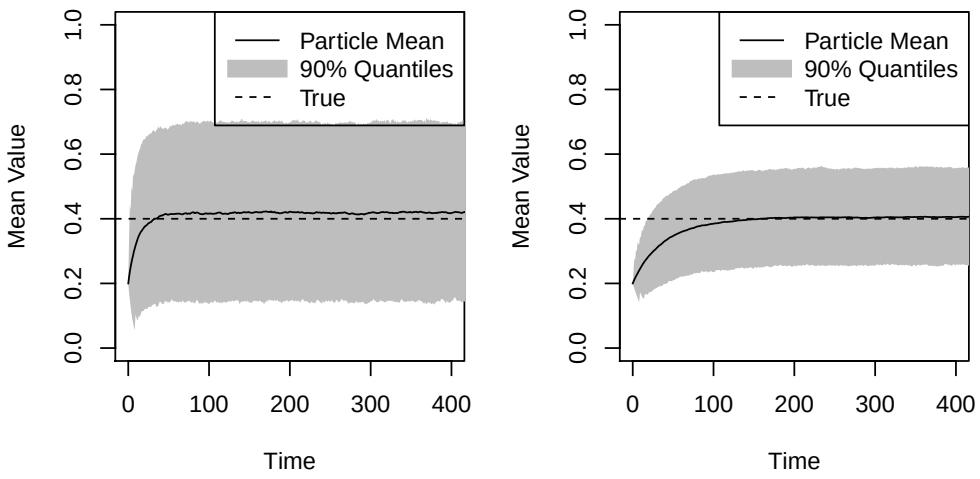
All the above follow via basic algebra manipulation and can be found in Appendix A.5.

Proof of Theorem 5.4.3. Lemmas 5.4.5, 5.4.6 and 5.4.7 prove that conditions 1,2 and 3 of Theorem 5.4.4 hold, which gives us that there exists at most one invariant measure.

To prove the existence of the invariant measure we need to satisfy the additional conditions 4 and 5. Since $P(x, \cdot)$ is a finite sum of atomic measures for every x in E, we get that P is Feller on E, satisfying condition 5. For condition 4 we look for a complete metric $d_0 \leq \sqrt{d}$. Indeed, since we can consider the process X to be defined on $[0, 1]$, we do this by extending the distance-like function d to include 0 and 1:

$$d(x, 1) = 1 \qquad d(1, 1) = 0 \qquad d(x, 0) = \min\left\{\left|\frac{1}{1-x} - 1\right|, 1\right\}.$$

Since for all $x, y \in [0, 1]$ we have that $d(x, y) \in [0, 1]$, we have that $\sqrt{d(x, y)} \geq d(x, y) = \frac{|x-y|}{(1-x)(1-y)} \geq |x - y|$. Setting d_0 to be the Euclidean metric, which is complete on $[0, 1]$ gives us the required condition. Therefore condition 4 of Theorem 5.4.4 is satisfied and hence there exists a unique invariant measure for the process X. □



(a) $N_1 = 6$, $N_2 = 6$ (b) $N_1 = 20$, $N_2 = 20$

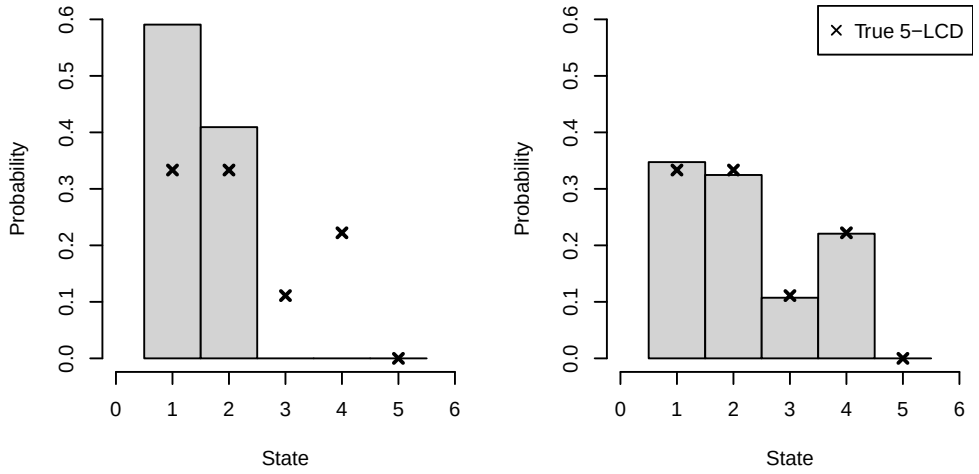
Figure 5.4.1: Mean convergence of 2-state Death process for different values of N_1, N_2:
$\delta = 0.4$, $T = 400$, Particles= 5000

Now we know that this Markov chain converges in a Wasserstein distance to some limiting distribution on $(0, 1)$, we would hope that the mean of the limiting distribution is close to the true value we want: δ. For linear systems this is simple to compute, but this is not the case for our process. However, one can see in Figure 5.4.1 that for even modest N_1, N_2 the process above does indeed converge to a value close to the true value δ, although larger values perform better. It is hoped that it may be possible to prove analytically that the mean of the distribution does in fact match the true value δ.

5.4.3 Applications: Death Process and Transient Immunity Process

Here we give some examples of scenarios where using the same number of particles and deterministic resampling scheme, we obtain better estimates for the LCD approximation through Regional Resampling. We first highlight the ability of the regional resampling scheme to repopulate regions of interest to avoid them becoming empty, a particular problem in reducible state spaces. Secondly, we illustrate its efficacy by demonstrating how the weight in different regions will converge.

To make these examples possible, we make use of O'Neill [2007] in which the author



(a) 1-Region Resampling (b) 2-Region Resampling

Figure 5.4.2: Comparison of simulations for 1-Region Resampling and 2-Region Resampling.
$\delta = (3, 2, 3, 1, 3)$, and $M = 100$, $\mathbf{N} = (40, 60)$, $T_{\text{end}} = 40$, $T_{\text{step}} = 1$, $T_b = 20$, $T_d = 2$

generates general birth-death processes with a pre-specified QSD, using the fact that QSDs are left eigenvectors of the transition rate matrix Q restricted to the transient state. Combining this with the fact that birth-death processes have tri-diagonal Q means we can easily generate death processes with whatever LCD we wish using

$$-\delta^* u_{i+1} = \delta_i u_i - \delta_{i+1} u_{i+1} \qquad i = 1, \ldots, L$$

In the case of the General Death process as described in Section 3.3, we consider the toy example on $S = \{1, \ldots, 5\}$ as described in Example 3.3.4 with the death rates given by $\delta = (\delta_1, \ldots, \delta_5) = (3, 2, 3, 1, 3)$. This has the true 5-LCD given by $\mathbf{u} = (1/3, 1/3, 1/9, 2/9, 0)$. We implement 2-Region Resampling by letting $S = S_1 \cup S_2$ with $S_1 = \{1, 2, 3\}$, and $S_2 = \{4, 5\}$.

Here we summarise the efficacy of this regional resampling in Figure 5.4.2 by noticing that in the 1-region case we only see particles in S_1, and hence the particles must fail to converge to the correct distribution. Indeed we make the following observation.

Proposition 5.4.8. *Let $(X(t))_{t\geq 0}$ be an absorbing Markov process on a reducible state space $S \cup \{0\}$ with initial distribution which gives weight to all communicating classes. Under the SMC sampler scheme, whenever a communicating class and all communicating classes from which it is accessible are empty, the SMC sampler scheme will instead converge to the $\mathbf{v}$-LCD where $\mathbf{v}$ gives weight only to those com-*

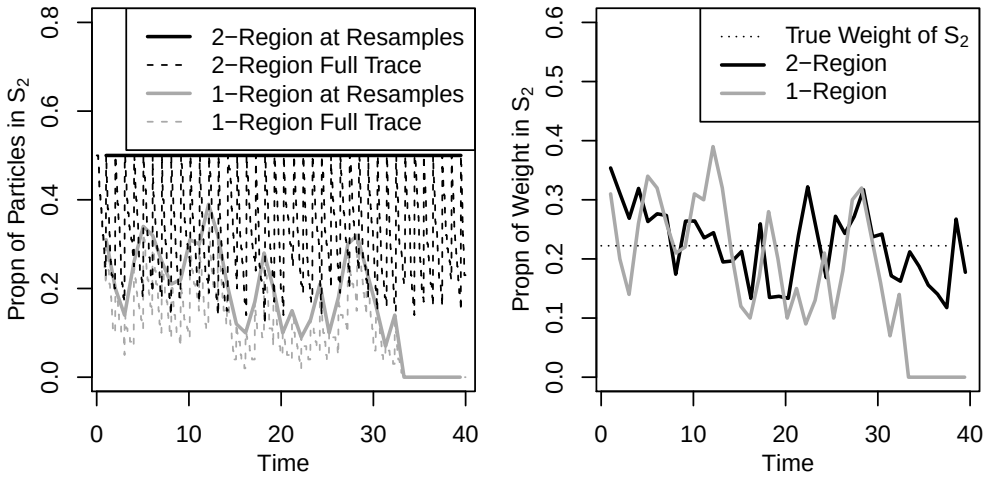
municating classes with particles remaining in them.

This can be seen by applying Theorem 2.1.11, and restarting the process whenever a communicating class becomes empty and cannot be refilled through evolution of the process; one can think of the process as evolving on the smaller state space excluding the empty classes. To illustrate it, we would note that for the example above, the SMC sampler scheme will start to converge to the 2-LCD as soon as states 4 and 5 are empty, and then converge (trivially) to the 1-LCD when state 2 is empty; there is only state 1 for the process to inhabit after this point.

On the other hand, one can see in Figure 5.4.2 the much better performance by adding just one extra region to the resampling method, and the resampling scheme does indeed reach good approximation of the true value of the LCD. Note that one can show that if the particles are all restricted to the regions closer to zero, then they will in fact begin to converge to the LCD with initial condition with mass only on the regions with particles still in them. Although 5 regions could have been specified to reflect the 5 transient communicating classes, in this example 2 regions were sufficient for the SMC sampler with regional resampling to converge to the correct distribution.

Figures 5.4.3a and 5.4.3b demonstrate how the proportion of particles and proportion of weights differ under particle refilling and regional resampling schemes. Although the number of particles following a resample in region S_2 is fixed for regional resampling, the proportion of weight in S_2 is not fixed, gradually converging to some value. Despite depletion between resampling times, at no point does the number of particles in S_2 reach zero, which ensures that regional resampling converges to the correct LCD.

Remark 5.4.9. One thing to choose in this resampling scheme is the division $(N_1, \ldots, N_L)$. This choice depends on the purpose of your simulation. For example, this chapter tries to explore the tails of LCDs. One way one could get more information on the tails is to simply choose N_l larger for regions S_l that you want more information on. However, in processes where the density of the LCD is high close to the absorbing state, the loss of weight at absorption times causes a large shift in the weighted mean of the particles. To counter this, one might choose to allocate a large number of particles to such region near to these maximal points. These options are neither mutually exclusive nor exhaustive, but merely illustrate that the choice of N_l can be done flexibly.



(a) Proportion of Particles in Region S_2 (b) Proportion of Weight in S_2

Figure 5.4.3: Comparison of weight and particle distribution for 1-Region Resampling and 2-Region Resampling.
$\delta = (3, 2, 3, 1, 3)$, and $M = 100$, $T_{\text{end}} = 40$, $T_d = 1$

Remark 5.4.10. We note here that, for the purposes of estimating LCDs, one should be aware of the distribution used to reallocate particles to locations. The use of proportional reallocation of particles to locations lends itself better to estimating mean behaviour. Using reallocation of particles to locations uniformly at random may lend itself better to tail estimation as we are more likely to allocate more particles to locations in the tails, but this comes with the drawback of having potentially larger weighted-particles close to the absorbing state, which disrupts convergence to the LCD. This is because if a large-weighted particle is absorbed, upon re normalization the weight in S is shifted significantly away from the locations close to absorption, and more time is needed to re-converge to the target distribution.

5.5 Stopping-Time Resampling Schemes

As mentioned in Section 2.3.3, there are three options for resampling timepoints which we can choose in our Regional Resampling. The simplest example is to make use of a deterministic timepoint scheme, where we perform resampling at given points of a timepoint sequence. One could use a random sequence of stopping times, independently chosen from the process. Alternatively, following work from Chen et al. [2005], Jenkins [2012], one can implement a stopping time resampling scheme such as the one we will now define.

One problem with the deterministic timepoint sequence used for L-Regional Resampling with reallocation using any of the resampling schemes mentioned in Section 2.3.3 or for Combine-Split is that run for long enough, one will encounter, with certainty, a period between two resampling timepoints during which all the particles leave one or more regions. At the extreme, any of the processes defined in Sections 2.2.1 3.1 and 4.1, if run for long enough will eventually experience a period between two resampling timepoints during which all the particles are absorbed. This is problematic since our resampling mechanisms may fail to function in these cases. Combine-Split requires particles to inhabit at least one non-absorbing location and Regional Resampling requires at least one particle in each region. For our process of interest, therefore, we introduce *L-Region Stopping-Time Regional Resampling* which triggers resampling precisely when any region has too few particles in it. We introduce a sequence of stopping times which activate resampling based on the diversity of particle locations. Other sequences of stopping times may be possible, along the lines of *multi-level SMC* as discussed in Chen et al. [2005]; Del Moral [2004]. In those algorithms, particles evolve independently between resampling events and stopping times are measurable with respect to the behaviour of any single particle.

Given a process $X = (X(t))_{t\geq 0}$ on the state space $S \cup \{0\}$ with transient states $S = \bigcup_{l=1}^{L} S_l$, we keep track of $M_l(t)$, the number of particles in region S_l at time t. Using this, our choice of $(N_1, \ldots, N_L)$ and a tuning parameter $\lambda \in [0, 1]$, we resample at the sequence of stopping times given by

$$T_k = \min\{T^{(1)}, \ldots, T^{(L)}, (T_{k-1} + T_{\max})\}$$

where the local stopping time for region S_l is given by

$$T^{(l)} = \inf\{t > T_{k-1} : M_l(t) < \lambda N_l\}$$

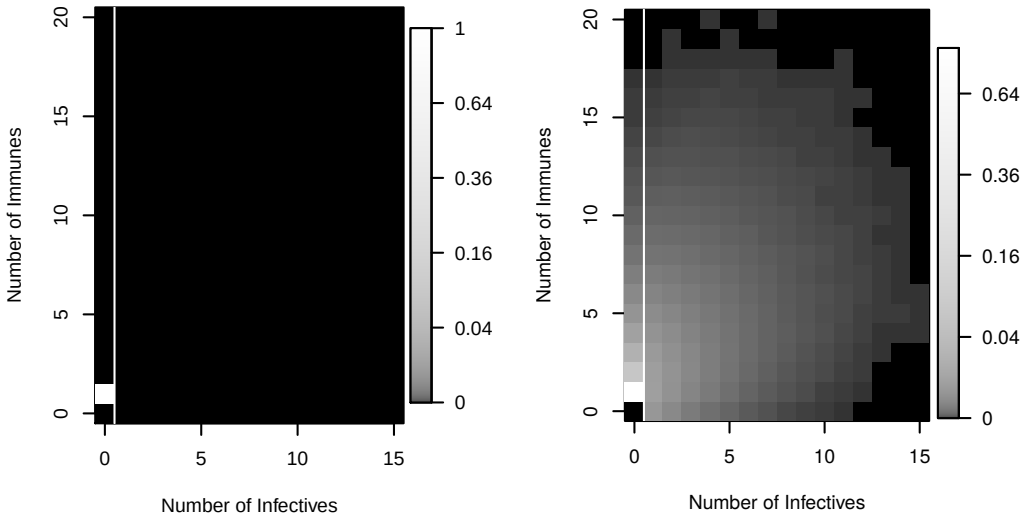
where $T_{\max} \in (0, \infty]$ is some fixed upper bound on resampling times. At each stopping time, we then perform L-Region Resampling. One should note that under L-Region Stopping Time Resampling, resamples happen at most $T_{\max}$ apart. To reduce the number of resamples, one increases $T_{\max}$ and allow the stochastic component of the stopping-times to trigger the resamples, rather than $T_{\max}$. Assuming that $N_l \geq 2$ for all l, then there must be at least one particle to resample from within each region at the time that a resampling event is triggered.

The main advantage to the Stopping-Time Resampling is that will definitely terminate successfully at the prescribed endpoint, whereas with all the deterministic

timepoint methods, there is the non-zero probability that during the simulation, a resampling step will be unable to take place. The second advantage, is also that one can tune λ, which dictates how low a region's population must get to trigger a resampling step. In doing so, one also tunes the number of resampling steps, and the speed of convergence in mean to the true LCD.

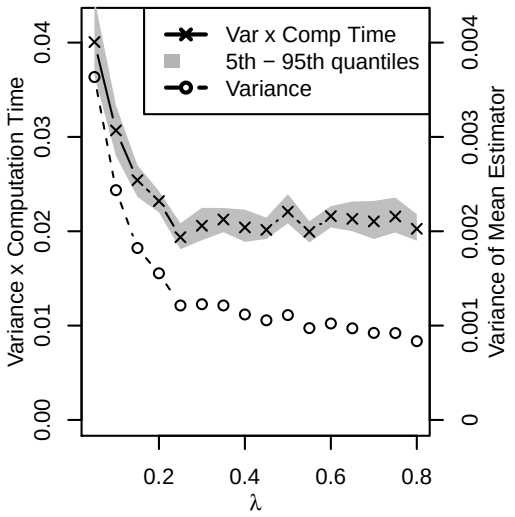
Demonstration: Transient Immunity Process

We return to the Transient Immunity Process to demonstrate the effects Regional Resampling with Stopping Times. The main point of Stopping-Time Regional Resampling is the guaranteed successful termination of the simulation, but we show how the tuning parameters can affect the performance of the simulation. We first show how 2-region resampling with dynamic resampling works for the Transient Immunity process to give an idea of the character of the $(1, 0)$-LCD as described in Section 4.3.1, making use the two regions $S_1 = \bigcup_{r=0}^{\infty}\{(0, r)\}$ and $S_2 = \{(i, r) : i > 0, r \geq 0\}$, so $S = S_1 \cup S_2$. In Figure 5.5.1a, we see that under 2-region resampling, despite having 40% of the particles in S_2 we see a negligible contribution from them. On the other hand, in Figure 5.5.1b, where the true $(1, 0)$-LCD gives weight to all states in the state space, the simulation gives a good visualisation of the $(1, 0)$-LCD.



(a) $\gamma - \beta = 0.5$, $\delta = 0.4$ (b) $\gamma - \beta = 0.5$, $\delta = 0.6$

Figure 5.5.1: Comparison of simulated LCDs from 2-region resampling: $M = 6000, \mathbf{N} = (4000, 2000)$, $T_{\text{end}} = 450$, $T_{\max} = 12$, $T_b = 40$, $T_d = 1$.



(a) Performance of estimator under varying λ: $T_{\max} = 10$, computed from 2000 runs.

λ	$T_{\max} = 1$	2	3	4	5
0	200	100	67	50	40
0.2	200	100	68	61	63
0.4	200	107	105	105	105
0.6	204	182	184	183	180
0.8	395	395	394	397	397

(b) Number of resampling steps as λ and $T_{\max}$ change

Figure 5.5.2: Comparison of performance under varying λ and $T_{\max}$:
$\beta - \gamma = 0.5$, $\delta = 0.6$, $T_{\text{end}} = 80$, $T_b = 0$, $T_d = 1$, $M = 500$, $\mathbf{N} = (300, 200)$

To illustrate how the choice of λ and $T_{\max}$ affect performance of the model, we consider how consistently such methods estimate the mean of the QSD for the Transient Immunity process. Under all choices of λ and $T_{\max}$ we obtained similar estimates of the mean, but the variability of such estimates is of interest. We first fix $T_{\max} = 10$. By repeating our simulations we measure the Monte Carlo variance of the estimator of the mean number of infectives, and the mean computation time to run a single simulation run of $M = 500$ particles; this computation time is only meant to be illustrative, and depends on the hardware and software used in general. From Figure 5.5.2a we see that the variance does indeed decrease as λ increases, but the improvement is noticeably smaller beyond $\lambda = 0.25$. If we take the product of the variance and mean computation time, then we see it is indeed optimal at this value of λ. For different uses, this optimal value of λ may vary, and if computation time is not an issue, one can instead choose as large a λ as possible.

If we furthermore vary $T_{\max}$ we see in Figure 5.5.2b that for different values of λ, and for sufficiently large $T_{\max}$, the precise choice of $T_{\max}$ is unimportant: there seems to develop a natural frequency of resampling steps. As such, it is felt that one should include a $T_{\max}$ which is slightly larger than the natural period of resampling under λ, since this $T_{\max}$ then still performs the job of avoiding particle weight degeneracy as in standard deterministic timepoint resampling schemes, but does not trigger

unnecessarily frequent resampling steps. For $\lambda = 0$ and $T_{\max} \geq 3$, the probability of a successful simulation run was low due to the high rate of particle extinction, and so repeated simulations were required.

Remark 5.5.1. Due to the fact that our stopping times depend on the states of all the particles at each time t, we cannot use the methods described in Jenkins [2012] to implement the stopping time resampling, and as such we need to simulate all particles simultaneously, which is only viable for processes such as the pure death process where at any given timepoint there are only a small number of subsequent events. Alternatively, one needs to throw away parts of each simulation if particles with a higher index trigger an earlier resampling. For the purposes of this demonstration, we just used a standard Gilespie algorithm.

5.6 Demonstration: Decay Estimation

We have developed a set of tools within the SMC Sampler framework, and demonstrated their efficacy. We now apply them to a known problem: evaluating the decay parameter or invariance value for LCDs associated to an absorbing process. In the finite state space, one could run into known issues with computing eigenvalues and eigenvectors for large non-sparse matrices. In the countable case, one needs to be able to evaluate $P_{ij}(t)$ and take limits to obtain expressions for α, which typically is difficult to compute if possible at all. Some work has been done in the case of general BDPs in van Doorn [2015] to obtain some representations of α, but these are of limited use in application. As previously discussed, the decay parameter can be found as the invariance value for processes starting from a single state. To this end, we apply the SMC sampler to obtain estimators for the invariance value x recalling Theorem 2.1.13, which states that if $\mathbf{u}$ is an x-invariant LCD, then $x = \sum_{s \in S} u_s q_{s0}$. If one can draw iid $X_j \sim \mathbf{u}$ for $j = 1, \ldots, M$ from the L-LCD then

$$\widehat{x} := \sum_{s \in S} q_{s0} \sum_{j=1}^{M} M^{-1} \mathbb{1}_{X^j = s}$$

is an unbiased estimator for the invariance value x. This estimator can therefore be implemented as follows. Following a burn-in period T_b, after which we would hope that the algorithm has mostly converged, then at each sampling timepoint t

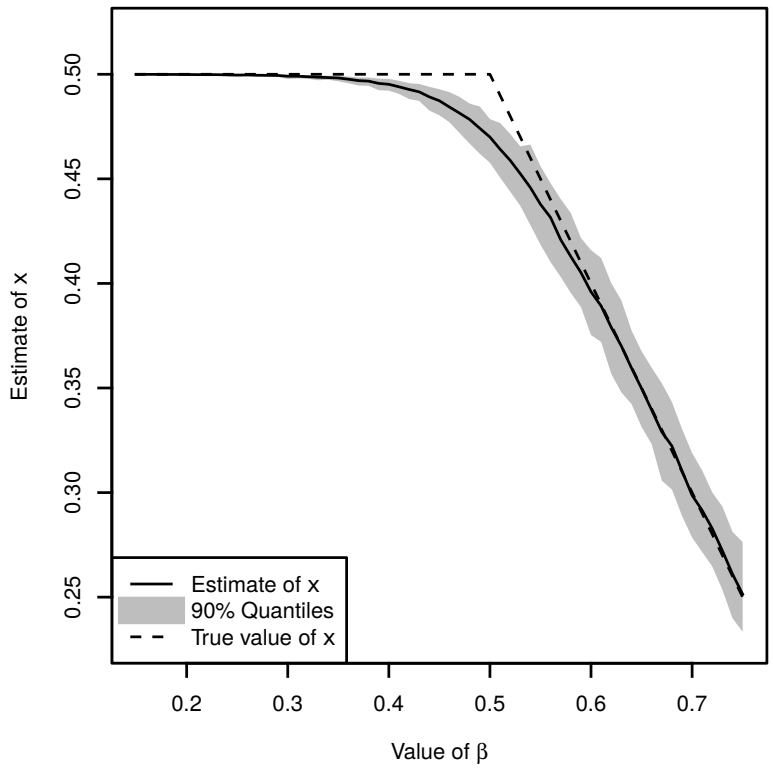


Figure 5.6.1: Estimate of x-invariance as β changes.
$\gamma = 1$, $\delta = 0.5$, $T_{\text{end}} = 60$, $T_{\max} = 5$, $T_b = 20$, $T_d = 1$, $M = 400$, $\mathbf{N} = (200, 200)$

we record

$$\widehat{x}_n := \sum_{s \in S} q_{s0} \sum_{j=1}^{M} W_j(t) \mathbb{1}_{X^j(t)=s}$$

where $\{W_j(t) : 1 \leq j \leq M\}$ are the normalised importance weights.

We apply this to the Transient Immunity Process. In this case, we consider the $(1,0)$-LCD, and try to estimate the invariance value x of the $(1,0)$-LCD conditional on $\{\mathbf{Y}(t) \neq (0,0)\}$. We note that since the only exit route is via $(I,R) = (0,1)$ we can define our estimator to be $\widehat{x} = \delta \sum_{j=1}^{M} W_j \mathbb{1}_{X^j=(0,1)}$. We have already shown in Theorem 4.3.5 that $x = \min(\delta, \gamma - \beta)$. To show this numerically, we make use of 2-region Stopping Time Regional Resampling with Combine-Split Particle Reallocation, and this is shown in Figure 5.6.1 for different values of β. The estimates confirm our expectations in the two regimes of the process, where δ is larger (or smaller) than $\gamma - \beta$. However, stochastic effects cause the invariance value to be underestimated close to the critical value $\delta = \gamma - \beta = 0.5$.

Remark 5.6.1. Since there is always a finite number of particles in the SMC sampler, it is not possible to use these techniques to sample from **v**-LCDs with initial distributions with infinite mean that sometimes exist for processes on infinite state spaces. Designing a mechanism to sample from such LCDs remains an open problem. To further this work, we would like to know whether there is a more systematic approach to the selection of λ and $\mathbf{N}$ to best approximate the true LCD.

Chapter 6

Evolving Strain Epidemic Models

6.1 Background

Following my work on the Transient Immunity Process in Chapter 4 and on the Linear BDP in Chapter 3, as well as pre-existing results on the SIS, and SIR stochastic epidemic models mentioned in Section 2.2, we look to construct a more realistic model which better describes the progression of pathogens such as seasonal influenza within a closed population over time. Standard epidemic models as discussed in Chapter 1 typically concern themselves with a single strain of a pathogen to which individuals are susceptible, become infectious with and then recover from. However, in pathogens such as seasonal influenza, there are often many different strains of the infection circulating within the population during a season, and as such each individual typically only has partial immunity to the collection of strains present. We will assume a linear progression of evolution of the virus, which is reasonable for seasonal Influenza A as demonstrated by the predominantly linear genetic evolution demonstrated in Fitch et al. [1997].

In the multi-strain models discussed in Neal [2006], there is assumed to be a finite number of possible strains, and each individual may be infected with one or more of such strains. Evolution has been modelled by a random jump process on the strain space $\{1, \ldots, K\}$. Gog and Grenfell [2002] uses a nearest neighbour jump process, where the strain an individual is infected with can jump to any adjacent strain. In

addition to this strain evolution is the notion of super-infectivity, where once an individual is infected with some strain, then their likelihood of being infected with a different strain during their initial infectious period is changed, either conferring extra susceptibility or immunity to other circulating strains. This is developed in Kamo and Sasaki [2002], but for our purposes, we will assume that each individual can only be infected with one strain of the infection at a time.

Where our model will differ from existing models is that we do not assume a finite number of possible strains. Instead, we will allow mutation of the pathogen such that each mutation event creates a previously unobserved strain of the pathogen. Models involving a countable number of infectious statuses have been discussed in the past Moy [1967], but these typically only use the previously mentioned nearest-neighbour evolution. In Moy [1967] this is expressed as a model for parasitic infections where the "type" of an individual is defined by the quantity of parasites in a host. It should be noted that within a finite population of size N, if each individual can only carry a single strain of the pathogen, then it must be the case that there are at most N unique strains within the population at any given time, but the specific strains may cycle and be replaced by newer ones. This can be compared to the BGW process discussed in Section 3.1, but modified to include mutations along the lifetimes of individuals; this is explored in Champagnat et al. [2012].

Despite this ever-progressing set of strains within the population, this chapter will focus on the number of currently circulating strains, and the differing proportions of immunity that the population possesses to each of these strains. To this end, we want to, in some sense, "fix" the current newest strain, and observe how the strain levels decay as we consider increasingly older strains.

In the rest of this chapter, we will define our main SIS-type and SIRS-type models; we will also define an equivalence relation to allow us to consider the relative strain immunity under the quasi-stationary distribution of the model. Under this equivalence relation, we also obtain a finite transient communicating class on which the model evolves, which according to Theorem 2.1.10 means that we have existence and uniqueness of the QSD. From this we will discuss the idea of marginals, and conclude some results regarding marginals of QSDs and LCDs. Using this we will obtain limiting results for extinction times of the models. Finally, we will look at features of the models' QSDs through a simulation study.

We will make use of properties of the evolving strain models which link them with

existing single-strain SIS and SIR models to allow us to make conclusions, but also have the detail of the full model to consider the notions of immunity within the population. Finally, we will consider the notion of strain diversity within a population.

6.2 Definition of the Model

Here we define the two models we are most interested in: the SIS and SIRS stochastic epidemic models with evolving strains that mutate during infection. We also define two infinite population models which are closely related: the Linear BDP and the Transient Immunity processes extended to include evolving strains.

6.2.1 SIS with Evolving Strains (E-SIS)

Consider a closed population of N individuals. We denote the total number of susceptibles by $\mathcal{S}(t)$ and the total number of infectives by $\mathcal{I}(t)$. Every individual will have a strain index $k \in \mathbb{Z}$ which denotes the most recent strain an individual was, or is currently, infected with. We denote the number of susceptibles with strain index k by $\mathcal{S}_k(t)$; infectives of strain index k are similarly denoted by $\mathcal{I}_k(t)$. Note $\mathcal{I}(t) = \sum_{k\in\mathbb{Z}} \mathcal{I}(t)$ and $\mathcal{S}(t) = \sum_{k\in\mathbb{Z}} \mathcal{S}_k(t)$. Finally we denote by $K^*(t)$ the maximal strain index observed up to time t and use $K^* = K^*(t^-)$ where the time is clear from context.

Strains are introduced into the population in the following way. We assume that during an infective's infectious period, the pathogen may mutate into previously unseen strains of the disease. As in the standard SIS model, we assume homogeneous mixing of individuals, and so each pair of individuals makes contact at the points of a Poisson process with rate $\beta > 0$. We also assume that all infectious periods are independent and identically distributed with $L \sim \text{Exp}(\gamma)$. Each time an infective makes contact with a susceptible individual, we assume that with some probability $\theta \in [0, 1]$, there is a successful infection of the susceptible with a previously unseen strain, which is given strain index $K^*(t^-) + 1$. With probability $1 - \theta$, the original strain in the infective attempts to infect the susceptible; the success of this infection depends on the strain index of the susceptible. For simplicity, we assume that immunity is cumulative: a susceptible with strain index k is immune to all infectives

with strain index $j \leq k$. We further assume all contacts, mutations, and infectious periods to be independent from each other.

Assuming some positive number of initial infectives $\mathcal{I}(0)$ and susceptibles $\mathcal{S}(0)$, the epidemic proceeds according to the following events.

Infection without Mutation:

$$(\mathcal{S}_j(t), \mathcal{I}_k(t)) \mapsto (\mathcal{S}_j(t) - 1, \mathcal{I}_k(t) + 1)$$

for $j < k$, with rate $\beta N^{-1}(1-\theta)\mathcal{S}_j(t)\mathcal{I}_k(t)$.

Infection with Mutation:

$$(\mathcal{S}_j(t), \mathcal{I}_{K^*+1}(t) = 0) \mapsto (\mathcal{S}_j(t) - 1, \mathcal{I}_{K^*+1}(t) = 1)$$

for $j \in \mathbb{N}$ with rate $\beta\theta N^{-1}\mathcal{S}_j(t)\mathcal{I}(t)$.

Recovery:

$$(\mathcal{S}_k(t), \mathcal{I}_k(t)) \mapsto (\mathcal{S}_k(t) + 1, \mathcal{I}_k(t) - 1)$$

for $k \in \mathbb{Z}$ with rate $\gamma\mathcal{I}_k(t)$.

We note that under the above events, we have that new strains emerge with rate $\beta\theta N^{-1}\mathcal{S}(t)\mathcal{I}(t)$.

6.2.2 SIRS with Evolving Strains (E-SIRS)

We will also consider a second model which introduces a transient period of global immunity following an infectious period. Each individual, following an infectious period, will experience a "globally immune" period, exponentially distributed amount of time with rate δ. After this they become susceptible, and still have the same strain index as their most recent infection. We will denote the number of immune individuals by $\mathcal{R}(t) = \sum_{k\in\mathbb{Z}} \mathcal{R}_k(t)$, where $\mathcal{R}_k(t)$ are the number of individuals who have just recovered from strain k but are still globally immune. The rest of the epidemic follows as in the SIS case. As before, we assume full independence of all contact, recovery and loss of immunity events, and assume a positive initial number of infectives $\mathcal{I}(0)$ and susceptibles $\mathcal{S}(0)$. To summarise the events:

Infection without Mutation:

$$(\mathcal{S}_j(t), \mathcal{I}_k(t)) \mapsto (\mathcal{S}_j(t) - 1, \mathcal{I}_k(t) - 1)$$

for $j < k$, with rate $\beta(1-\theta)N^{-1}\mathcal{S}_j(t)\mathcal{I}_k(t)$.

Infection with Mutation:

$$(\mathcal{S}_j(t), \mathcal{I}_{K^*+1}(t) = 0) \mapsto (\mathcal{S}_j(t) - 1, \mathcal{I}_{K^*+1}(t) = 1)$$

for $j \in \mathbb{Z}$ with rate $\beta\theta N^{-1}\mathcal{S}_j(t)\mathcal{I}(t)$

Recovery:

$$(\mathcal{I}_k(t), \mathcal{R}_k(t)) \mapsto (\mathcal{I}_k(t) - 1, \mathcal{R}_k(t) + 1)$$

for $k \in \mathbb{Z}$ with rate $\gamma\mathcal{I}_k(t)$

Loss of Global Immunity:

$$(\mathcal{R}_k(t), \mathcal{S}_k(t)) \mapsto (\mathcal{R}_k(t) - 1, \mathcal{S}_k(t) + 1)$$

for $k \in \mathbb{Z}$ with rate $\delta\mathcal{R}_k(t)$

Here we use the same notation (such as $\mathcal{S}_k(t)$) as for the E-SIS model in Subsection 6.2.1, as it will always be clear from the context which model we are referring to. We also note at this point that, under the above events, mutations occur with rate $\beta\theta N^{-1}\mathcal{S}(t)\mathcal{I}(t)$, as in the SIS Evolving strain model.

6.2.3 Linear BDP with Evolving Strains (E-BDP)

In order to obtain some limiting results for extinction times, we will also consider two "infinite" population models.

This first limiting model corresponds to the Evolving SIS model. If we assume there to be an infinite number of initial susceptibles that are not immune to any strains at the start of the epidemic, then we can assume the infections will always be successful. We denote the total number of infectives by $\mathcal{I}(t)$ and the number of infectives of strain k by $\mathcal{I}_k(t)$. Here we assume infections from a given strain k and recoveries from that strain to behave as a linear birth-death process with birth rate

β and death rate γ. Additionally, at the point of each infection, with probability $\theta \in [0, 1]$, the new infective is infected with a previously unseen strain, and given the relevant strain index $k = K^* + 1$. To summarise the events:

Infection without Mutation:

$$\mathcal{I}_k(t) \mapsto \mathcal{I}_k(t) + 1$$

for $k \in \mathbb{Z}$, with rate $\beta(1-\theta)\mathcal{I}_k(t)$

Infection with Mutation:

$$\mathcal{I}_{K^*+1}(t) = 0 \mapsto \mathcal{I}_{K^*+1}(t) = 1$$

with rate $\beta\theta\mathcal{I}(t)$

Recovery:

$$\mathcal{I}_k(t) \mapsto \mathcal{I}_k(t) - 1$$

for $k \in \mathbb{Z}$ with rate $\gamma\mathcal{I}_k(t)$

For this model, we see mutations occur with rate $\beta\theta\mathcal{I}(t)$. Moreover, it should be noticed that each strain behaves, conditional on its time of emergence, according to a linear birth-death process with birth rate $\beta(1-\theta)$ and death rate γ. The total number of infectives $\mathcal{I}(t)$ also behaves according a birth-death process with birth rate β and death rate γ.

6.2.4 Transient Immunity Process with Evolving Strains (E-TI)

Our final model of interest is the infinite population of the Evolving SIRS model in Section 6.2.2. As in the single strain case, we consider the Transient Immunity process as in Chapter 4, and add a mutation mechanism. In this case, we assume an infinite population of susceptibles as in the Evolving Linear BDP. Infections happen as in Section 6.2.4 but following infections, individuals go into a globally immune period before becoming susceptible again. Mutations occur as in Subsection 6.2.1, as do the strain indexing conventions. To summarise the events:

Infection without Mutation:

$$\mathcal{I}_k(t) \mapsto \mathcal{I}_k(t) + 1$$

with rate $\beta(1-\theta)\mathcal{I}_k(t)$

Infection with Mutation:

$$\mathcal{I}_{K^*+1}(t) = 0 \mapsto \mathcal{I}_{K^*+1}(t) = 1$$

with rate $\beta\theta\mathcal{I}(t)$

Recovery:

$$(\mathcal{I}_k(t), \mathcal{R}_k(t)) \mapsto (\mathcal{I}_k(t) - 1, \mathcal{R}_k(t) + 1)$$

with rate $\gamma\mathcal{I}_k(t)$

Loss of Global Immunity:

$$\mathcal{R}_k(t) \mapsto \mathcal{R}_k(t) - 1$$

with rate $\delta\mathcal{R}_k(t)$

6.2.5 State Space

We can describe the state space for each of the Evolving Strain models as follows:

- E-SIS model:

  $$\Omega^{(1)}_{E-SIS} = \{(\mathbf{S}, \mathbf{I}) : \sum_{k\in\mathbb{Z}} (S_k + I_k) = N\}$$

  with $\mathbf{S}, \mathbf{I}$ being infinite sequences taking values in $\{0, \ldots, N\}$.

- E-SIRS model:

  $$\Omega^{(1)}_{E-SIRS} = \{(\mathbf{S}, \mathbf{I}, \mathbf{R}) : \sum_{k\in\mathbb{Z}} (S_k + I_k + R_k) = N\}$$

  with $\mathbf{S}, \mathbf{I}, \mathbf{R}$ being infinite sequences taking values in $\{0, \ldots, N\}$.

- E-BDP:

$$\Omega^{(1)}_{E-BDP} = \{\mathbf{I} : I_k \in \mathbb{N}_0 \text{ for all } k \in \mathbb{Z}\}$$

- E-TI process:

$$\Omega^{(1)}_{E-TI} = \{(\mathbf{I}, \mathbf{R}) : I_k, R_k \in \mathbb{N}_0 \text{ for all } k \in \mathbb{Z}\}$$

For generalisation, we will often suppress the subscripts when the context is clear, so Ω can refer to any of the above. We will also use the wider space $\bar{\Omega}^{(1)} = \{(\mathbf{S}, \mathbf{I}, \mathbf{R}) : S_k, I_k, R_k \in \mathbb{N}_0 \ \forall k \in \mathbb{Z}\}$ for some arguments.

6.2.6 Linking to Single-Strain Models

In this section we make two main observations which link the Evolving Strain models defined in Section 6.2 to the single-strain models defined in Section 2.2.

Consider the E-SIS model starting with some number of initial infectives of strain 0 and susceptibles of strain index -1. If we set $\theta = 0$, then this means that mutation is impossible. As a result, once an individual has become infected and recovered, they join the S_0 class. However, since mutation is impossible, all infection attempts will be unsuccessful. In this way S_0 behaves in an identical fashion to the R class in the SIR model, and we can identify the two models in this way.

Similarly, consider the same model but with $\theta = 1$. In this case, all contacts are mutation contacts, so all susceptible individuals will be successfully infected upon contact with an infective. Moreover, strain indices become irrelevant since there is no immunity in the population. In this case, we can identify the total number of susceptibles in the E-SIS model with the number of susceptibles in the single-strain SIS model, and similarly for the number of infectives. With these, we can identify the E-SIS model with $\theta = 1$ with the single strain SIS model.

For $\theta = 1$ we can also perform a similar identification between the SIRS model and the E-SIRS model, using the total loss of immunity described in the previous paragraph.

6.3 Equivalence Relation

With the current models, we observe that the state space $\bar{\Omega}^{(1)}$ is uncountably infinite, looking like $\mathbb{N}^{3\mathbb{N}}$, and due to the evolution of the strains in the model, we cannot ensure that the processes have LCDs or QSDs. Indeed, one might expect non-existence, since we would observe the number of emerged strains $K^*(t)$ to become infinite conditional on the event $\{\mathcal{I}(t) > 0\}$. To this end, and to better observe the expected number of strains in circulation, we implement an equivalence relation on the state space. Firstly we need to introduce some notation.

Definition 6.3.1. The *active strain set* of a state $(\mathbf{S}, \mathbf{I}, \mathbf{R}) \in \bar{\Omega}^{(1)}$ is given by $\mathbb{K} = \{k \in \mathbb{Z} : I_k > 0\}$. The elements of this set are indexed over $\mathbb{N}$ in ascending order, so for $k_a, k_b \in \mathbb{K}$, we have $k_a < k_b$ whenever $0 < a < b$.

Using this notation, we define the following equivalence relation on $\bar{\Omega}^{(1)}$, and subsequently all the spaces defined in Section 6.2.5.

Definition 6.3.2. Two states $(\mathbf{S}, \mathbf{I}, \mathbf{R})$ and $(\mathbf{S}', \mathbf{I}', \mathbf{R}') \in \bar{\Omega}^{(1)}$ are equivalent if and only if the following conditions hold.

- $\mathcal{S} = \sum_{k\in\mathbb{K}} \mathcal{S}_k = \sum_{k'\in\mathbb{K}'} \mathcal{S}'_k = \mathcal{S}'$, and similarly $\mathcal{I} = \mathcal{I}'$ and $\mathcal{R} = \mathcal{R}'$, so the number of susceptibles, infectives and immunes match.
- There are the same number of active strains in the population: $|\mathbb{K}| = |\mathbb{K}'|$.
- $\sum_{k<k_a} \mathcal{S}_k = \sum_{k<k'_a} \mathcal{S}'_k \qquad$ for $a = 1, \ldots, |\mathbb{K}|$
- $\mathcal{I}_{k_a} = \mathcal{I}'_{k'_a} \qquad$ for $a = 1, \ldots, |\mathbb{K}|$
- $\sum_{k<k_a} \mathcal{R}_k = \sum_{k<k'_a} \mathcal{R}'_k \qquad$ for $a = 1, \ldots, |\mathbb{K}|$

What this equivalence relation does is it considers states to be equivalent if the numbers of infectives in each active strain are the same, and the number of immunes and susceptibles with indices between two active strains are the same. In order to easily refer to the equivalence classes, we define the following representative of each equivalence class.

Definition 6.3.3. The *representative* of the equivalence class containing $(\mathbf{S}, \mathbf{I}, \mathbf{R})$ is $(\mathbf{S}^*, \mathbf{I}^*, \mathbf{R}^*)$ where the corresponding active strains for the representative are given

by $\mathbb{K}^* = \{1 - |\mathbb{K}|, \dots, 0\}$. More specifically if we let $\phi : \mathbb{K} \to \mathbb{K}^*$ be defined by $\phi(k_a) = a - |\mathbb{K}|$ for $a = 1 \dots, \mathbb{K}$ then $(\mathbf{S}^*, \mathbf{I}^*, \mathbf{R}^*)$ is given by

$$I_j^* = \begin{cases} I_k & \text{for } j = \phi(k), k \in \mathbb{K} \\ 0 & \text{otherwise} \end{cases}$$

with

$$S_k^* = \sum_{j=\phi^{-1}(k)}^{\phi^{-1}(k+1)-1} S_j \qquad R_k^* = \sum_{j=\phi^{-1}(k)}^{\phi^{-1}(k+1)-1} R_j, \qquad \text{for } k \in \mathbb{K}^* \setminus \{0, |\mathbb{K}| - 1\}$$

$$S_0^* = \sum_{j=\phi^{-1}(0)}^{\infty} S_j \qquad R_0^* = \sum_{j \geq \phi^{-1}(0)} R_j$$

$$S_{1-|\mathbb{K}|}^* = \sum_{j=-\infty}^{\phi^{-1}(1-|\mathbb{K}|)-1} S_j \quad R_{1-|\mathbb{K}|}^* = \sum_{j=-\infty}^{\phi^{-1}(1-|\mathbb{K}|)-1} R_j$$

What this definition does is remove all empty strains (strains with no infective individuals with that index), and makes the most recent strain have the index 0. All susceptibles and removed individuals are given the strain index one less than the nearest infective above them in strain order. Any totally immune susceptibles are given strain 0, as though they just recovered from the most recently emerged strain.

This is more easily explained in the following examples.

Example 6.3.4. Consider the state $(\mathbf{S}, \mathbf{I}, \mathbf{R}) \in \Omega^{(1)}$ illustrated in Figure 6.3.1a and given by

$$(\dots, S_1, \dots, S_7, \dots) = (\dots, 0, 0, 1, 0, 0, 0, 1, \dots)$$
$$(\dots, I_1, \dots, I_7, \dots) = (\dots, 0, 1, 0, 0, 0, 1, 0, \dots)$$
$$(\dots, R_1, \dots, R_7, \dots) = (\dots, 1, 0, 0, 0, 1, 0, 0, \dots)$$

where all omitted terms of $\mathbf{S}$, $\mathbf{I}$, $\mathbf{R}$ are given to be zero. Then we see that $\mathbb{K} = \{2, 6\} \Rightarrow \mathbb{K}^* = \{-1, 0\}$ with $\phi(2) = -1, \phi(6) = 0$ and the representative under the

equivalence relation is illustrated in Figure 6.3.1b, and is given by

$$(S_{-2}, \dots, S_0) = (0, 1, 1)$$
$$(I_{-2}, \dots, I_0) = (0, 1, 1)$$
$$(R_{-2}, \dots, R_0) = (1, 1, 0)$$

R	I	S		R	I	S
1	2	3	4	5	6	7

(a) Example Population 1

		S
	I	S
R	R	I
-2	-1	0

(b) Representative of Population 1

Figure 6.3.1: Example of Equivalence Representative for Population 1

Example 6.3.5. Consider the state $(\mathbf{S}, \mathbf{I}, \mathbf{R}) \in \Omega^{(1)}$ illustrated in Figure 6.3.2a and given by

$$(\dots, S_1, \dots, S_6, \dots) = (\dots, 1, 2, 0, 0, 2, 1, \dots)$$
$$(\dots, I_1, \dots, I_6, \dots) = (\dots, 0, 1, 0, 2, 3, 0, \dots)$$
$$(\dots, R_1, \dots, R_6, \dots) = (\dots, 1, 1, 0, 2, 1, 1, \dots)$$

where all other terms of $\mathbf{S}$, $\mathbf{I}$ and $\mathbf{R}$ are zero. With this $\mathbb{K} = \{2, 4, 5\} \Rightarrow \mathbb{K}^* = \{-2, -1, 0\}$ and so the representative under the equivalence relation is illustrated in Figure 6.3.2b given by

$$(S_{-3}, \dots, S_0) = (1, 2, 0, 3)$$
$$(I_{-3}, \dots, I_0) = (0, 1, 2, 3)$$
$$(R_{-3}, \dots, R_0) = (1, 1, 2, 2)$$

6.3.1 Notation for processes

Here we make clear the notation for what follows. Without the equivalence relation, we denote the state space of the relevant process by $\Omega^{(1)}$; subscripts will be used if ambiguity arises. With the equivalence relation, we denote the state space of the process under this equivalence by Ω, again with the relevant subscripts where needed. We will identify all the absorbing states to the state $\mathbf{0} = (0, 0, 0)$. We will use $\mathbf{x} = (\mathbf{S}, \mathbf{I}, \mathbf{R})$, with, for example, $\mathbf{S} = (S_k)_k \in \mathbb{N}_0^{\mathbb{Z}}$, to denote a typical element of

				S	
				S	
	S		I	I	
	S		I	I	
S	I		R	I	S
I	R		R	R	R
1	2	3	4	5	6

(a) Example Population 2

			S
			S
			S
			I
	S	I	I
	S	I	I
S	I	R	R
I	R	R	R
-3	-2	-1	0

(b) Representative of Population 2

Figure 6.3.2: Example of Representative for Population 2

the state space. We will also use, for example, $|\mathbf{S}| = \sum_{k\in\mathbb{Z}} S_k$ to denote the number of susceptibles.

For any of the models defined in Section 6.2, a statistic or random variable written in calligraphic type, e.g. $\mathcal{R}_k(t)$ refers to a process without the equivalence relation. The corresponding statistics and variables written in roman type e.g. $R_k(t)$ refer to the processes under the equivalence relation.

6.4 Quasi-Stationarity and Absorbing States

For the models defined in Section 6.2, we are interested in which states correspond to absorption and to the end of the epidemic. We note that since every susceptible-infective contact can result in an infection due to the mutation mechanism (assuming $\theta > 0$), an absorbing subset of the state space is reached only when $I(t) = 0$. In some models, such as the E-SIRS model, there are also transient subsets within this set.

6.4.1 Existence and Uniqueness

Here we will summarise the existence and uniqueness results for the processes defined in Section 6.2.

Theorem 6.4.1. *Let the E-SIS model be defined as in Subsection 6.2.1 with parameters $\beta, \gamma, N > 0$ and $\theta \in (0, 1]$. Then, conditional on the event $\{I(t) > 0\}$ there exists a unique QSD which equals the unique LCD of the process, and gives weight*

to all states. If $\theta = 0$, then there exists a unique QSD which equals the unique LCD of the process which gives full weight to the state with a single infective with strain index 0, and $N-1$ susceptibles with strain index 0.

Proof. For $\theta \in (0, 1]$, we obtain existence and uniqueness by proving that S is a single finite communicating class, which immediately gives existence and uniqueness by van Doorn and Pollett [2013].

We note that, under the equivalence relation, there can be at most N different strain indices present in the population. Therefore we can bound the size of S, the set of transient states, by N^{2N}. Indeed, since we need only consider the values $(S_{1-N}, \ldots, S_0)$, and $(I_{1-N}, \ldots, I_0)$, there are only $2N$ dimensions to the process, and each can only take values in $\{0, \ldots, N\}$.

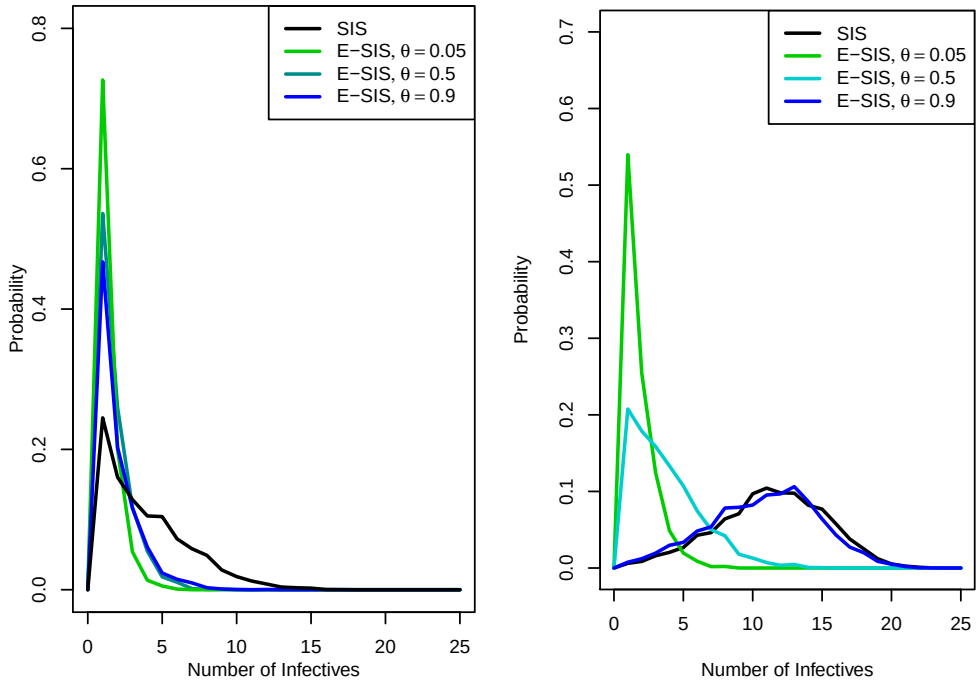
One can see that the transient states form a single communicating class by noting that one can get from a single infective of strain index 0 with $N-1$ susceptibles of index 0 to any other state through infections (mutation and non-mutation) and recoveries, and if all but one infective recovers, then the process can return to the single infective case mentioned above.

For $\theta = 0$ we recall Section 6.2.6 and note that we can identify the E-SIS and SIR models in this case. The theorem then follows by directly applying Proposition 2.2.2 converting the "removed" individuals in the SIR model to susceptibles of strain index 0 in the E-SIS model. □

In Figure 6.4.1 we see the number of infectives under different values of β and θ and note that in both the subcritical case $\beta = 0.5$ and the supercritical case $\beta = 2$ we see that the distribution for the $E-SIS$ model is skewed more towards fewer infectives for lower values of θ, and gets closer to the SIS model as $\theta \to 1$. These, along with all the figures in this chapter were generated using the techniques in Chapter 5.

Theorem 6.4.2. *Let the E-SIRS model be defined as in Subsection 6.2.2 with parameters $\beta, \gamma, \delta, N > 0$ and $\theta \in (0, 1]$. Then, conditional on $\{I(t) > 0\}$ there exists a unique QSD which equals the unique LCD of the process. If $\theta = 0$, this QSD still exists, and gives full weight to the state with one infective with strain index 0 and $N-1$ susceptibles with strain index 0.*

Proof. For the $\theta \in (0, 1]$, one follows the same argument as in Theorem 6.4.1, this

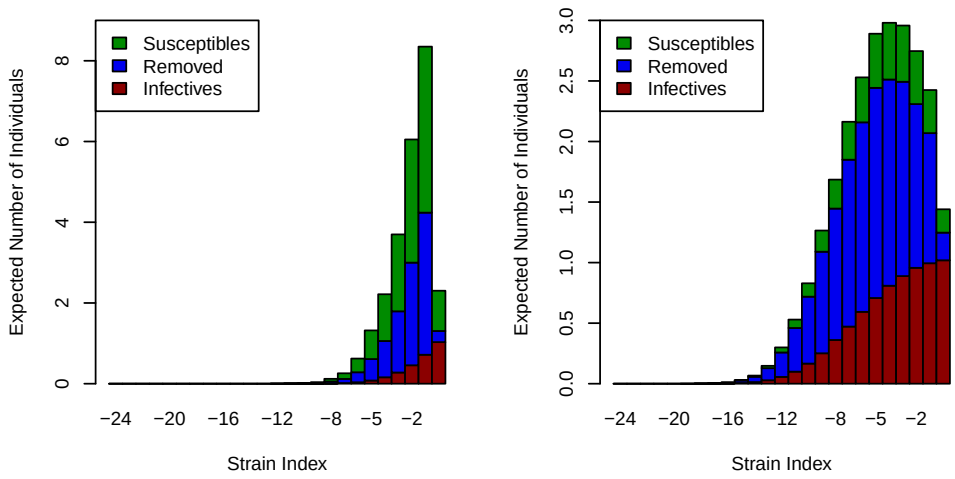


(a) $\beta = 0.5$, $\gamma = 1$, $N = 25$ (b) $\beta = 2$, $\gamma = 1$, $N = 25$

Figure 6.4.1: Number of infectives under QSD for the SIS and E-SIS Models under different values of β, θ. Simulations produced using SMC Sampler:
$M = 400$, $T_{\max} = 20$, $T_b = 100$, $T_d = 0.25$, $T_{\text{end}} = 140$, $\lambda = 0.4$

time bounding the size of the state space by N^{3N}, since we also need to consider $(R_{1-N}, \ldots, R_0)$. The fact that the transient states form a single communicating class also follows as in Theorem 6.4.1. For $\theta = 0$, we see that each state that can be reached is a transient communicating class; there is no way to return to a state once left. As such, we need to consider the decay parameter on leaving each state, which equals the exponential rate of leaving such a state. This equals $\beta S_k I_k/N + \gamma I_k + \delta R_k$. The decay parameter for the process is therefore the minimal such value across all non-absorbing states. We can therefore choose $S_{-1} = N - 1$, $I_0 = 1$ and zero otherwise to minimise the decay parameter. According to van Doorn and Pollett [2008] this forces the QSD to have full mass on this state where $I_0 = 1, S_{-1} = N - 1$ and all other values are zero, since the only states accessible from this state are absorbing ones. □

In Figure 6.4.2, we see examples of the mean under the QSD and see that in the case of the lower value of δ there are a larger proportion of "Removed" individuals which are globally immune, as expected, causing a lesser number of strains in the



(a) $\beta = 2$, $\theta = 0.9$, $\delta = 0.2$,$N = 25$ (b) $\beta = 2$, $\theta = 0.9$, $\delta = 2$, $N = 25$

Figure 6.4.2: Comparisons of expected population make-up under E-SIRS QSDs. Simulations produced using SMC Sampler:
$M = 400$, $T_{\max} = 20$, $T_b = 100$, $T_d = 0.25$, $T_{\text{end}} = 140$, $\lambda = 0.4$

mean of the epidemic.

Theorem 6.4.3. *Let $\mathbf{I}(t)$ be the E-BDP with parameters $\gamma > \beta > 0$ and $\theta \in [0, 1]$. Then, conditional on the event $\{I(t) > 0\}$, there exists a unique α-invariant QSD, where $\alpha = \gamma - \beta$.*

Proof. To prove existence, we first show that the state space we are interested in is countable. To do this we use the following construction. Starting with a single infective of strain 0, we can define a method of constructing the state space. By having a birth in strain 0, or a mutation event, one can systematically arrive at any state in the state space. Given these two possible events, one can encode each state according to a finite binary sequence, which corresponds to a unique integer which we can use to enumerate the space. Given that there exists a lower bound $l \in \mathbb{Z}$ such that $I_k = 0$ for all $k \leq l$, we construct the state as follows. Starting with the lowest non-zero strain index $l+1$ consider I_{l+1} strain 0 within-strain-infection events. Then for each higher strain k, we choose a mutation event followed by $I_k - 1$ within-strain infection events. See Figure 6.4.3 for an example of how this works. Note that only considering finite sequences gives countability, unlike the uncountability of the infinite paths on this binary tree.

To obtain existence of a QSD, we now introduce a coupling. Let $\mathbf{I}(t) = (X_j(t))_{j \in \mathbb{Z}}$ be the E-BDP defined on a probability space $(E, \mathcal{F}, \mathbb{P})$, so $\mathbf{I} : [0, \infty) \times E \to \Omega_{E=BDP}$.

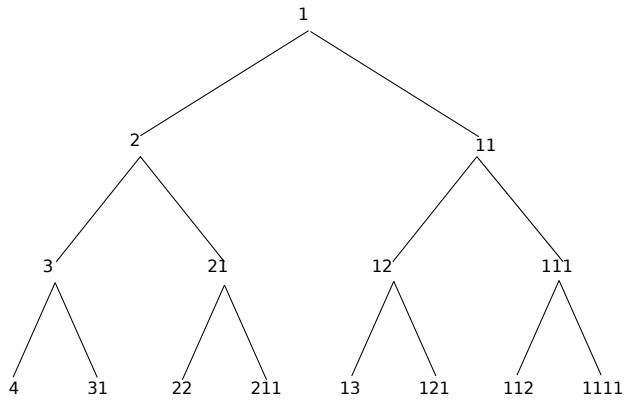


Figure 6.4.3: Systematic Generation of State Space for the E-BDP. The label node on the tree above corresponds to the state $\mathbf{I} = (..., I_{-3}, I_{-2}, I_{-1}, I_0)$ where all omitted labels are zero.

Let α^X be the decay parameter for $\mathbf{I}(t)$. Let $Y(t)$ be the process defined on the same probability space, given by $Y(t) = \sum_{j \in Z} X_j(t)$. Since the mutations do not affect whether infections are successful or not, $Y(t)$ can be seen to be a single-strain linear BDP with birth rate β, and death rate γ. As discussed in Section 3.2, $Y(t)$ has the decay parameter $\alpha^Y = \gamma - \beta$. Let T_X be the extinction time of $\mathbf{I}(t)$ and T_Y for $Y(t)$.

Letting α_0^Y be the absorption parameter for $Y(t)$, and α_0^X for $\mathbf{I}(t)$, we also know that $\alpha_0^Y = \gamma - \beta$ also. Since $T_X = T_Y$, we use the definition of the absorption parameter to deduce that $\alpha_0^X = \gamma - \beta$, and hence $\alpha^X \geq \alpha_0^X > 0$. Using Theorem 2.1.17 we get existence of a QSD. Moreover, using Theorem 2.1.14, we must have $\alpha_X = \gamma - \beta$ since there is only one state from which extinction can occur: one must have 1 infective before extinction, which must be of strain 0 by the equivalence relation. This also leads to the uniqueness of the α-invariant QSD. □

Here we develop some results on the marginals of the QSDs for the different processes in a similar fashion to that of Proposition 4.2.3, relating LCDs and QSDs of the more complex evolving strain models to their single-strain counterparts, and relating the marginals of QSDs of the Evolving Transient Immunity process to QSDs of the E-BDP.

Theorem 6.4.4. *Let* $\mathbf{u} = (u_{\mathbf{x}})_{\mathbf{x} \in \Omega_{E-BDP} \setminus \{\mathbf{0}\}}$ *be a* $\mathbf{y}$*-LCD for the E-BDP and let* $\mathbf{u}^I = (u_j^I)_j \geq 0$ *be defined by*

$$u_j^I = \sum_{\mathbf{x}:|\mathbf{I}|=j} u_{\mathbf{x}} \tag{6.4.1}$$

Then the i-LCD for the Linear BDP is given by $|\{\mathbf{x} : |\mathbf{x}| = i\}|\mathbf{u}^I$.

Proof. Fix $i > 0$. Under the equivalence relation, we know that i individuals can inhabit at most i strains which must be strains $\{1-i, \dots, 0\}$ for the equivalence representative. This means that $|\{\mathbf{x} : |\mathbf{x}| = i\}|$ is finite, and so (6.4.1) is well defined. Now, we create a coupling between the Linear BDP $Y(t)$ and the E-BDP $\mathbf{X}(t)$ on a common probability space $(E, \mathcal{F}, \mathbb{P})$. In this case we define $Y(t, \omega) = |\mathbf{X}(t, \omega)|$. Using this coupling and Definition 2.1.1 we note that we can decompose events for the Y-process into a disjoint union of events for the $\mathbf{X}$-process:

$$\{Y(t) = j, Y(0) = k\} = \bigcup_{|\mathbf{x}|=j} \bigcup_{|\mathbf{y}|=k} \{\mathbf{X}(t) = \mathbf{x}, \mathbf{X}(0) = \mathbf{y}\}$$
$$\{Y(t) > 0\} = \{\mathbf{X}(t) \neq \mathbf{0}\}$$

Similarly to Theorem 4.2.3, we let $\mathcal{F}_t = \sigma(\{\mathbf{X}(s) : s \leq t\})$ and $\mathcal{G}_t = \sigma(\{Y(s) : s \leq t\})$. Under these we get by construction that $\mathbb{P}[Y(t) \in A|\mathcal{F}_t] = \mathbb{P}[Y(t) \in A|\mathcal{G}_t]$. With this we see that

$$\lim_{t\to\infty} \mathbb{P}_i[Y(t) = j | Y(t) > 0] = \lim_{t\to\infty} \sum_{|\mathbf{x}|=i} \sum_{|\mathbf{y}|=j} \mathbb{P}_{\mathbf{x}}[\mathbf{X}(t) = \mathbf{y} | \mathbf{X}(t) \neq \mathbf{0}]$$

Since this is a well-defined probability for each $t \geq 0$, $i, j \geq 1$ we can use Dominated Convergence to swap the limit and the sums to obtain

$$\begin{aligned}\lim_{t\to\infty} \mathbb{P}_i[Y(t) = j | Y(t) > 0] &= \sum_{|\mathbf{x}|=i} \sum_{|\mathbf{y}|=j} \lim_{t\to\infty} \mathbb{P}_{\mathbf{x}}[\mathbf{X}(t) = \mathbf{y} | \mathbf{X}(t) \neq \mathbf{0}] \\ &= \sum_{|\mathbf{x}|=i} \sum_{|\mathbf{y}|=j} u_{\mathbf{y}} = \sum_{|\mathbf{x}|=i} u_j^I = |\{|\mathbf{x}| = i\}| u_j^I\end{aligned}$$

Where $\mathbf{u} = (u_{\mathbf{y}})_{\mathbf{y}}$ is the common $\mathbf{x}$-LCD for each $\mathbf{x} \in S$; this uniqueness follows from the fact that the transient states form a single communicating class, and existence of such LCDs was proven in the Section 6.4.1. From the independence of $\mathbf{x}$ and the fact that the sum in $\mathbf{x}$ is finite under the equivalence relation, we get that $\sum_{|\mathbf{x}|=i} u_j^I = |\{|\mathbf{x}| = i\}| u_j^I$ and the result is shown. □

6.4.2 Inflated Mutation Rates under Conditioning

With regards to mutation mechanism, we expect that under conditioning to see a greater number of mutation events than without the conditioning.

To see this in effect for the E-SIS model, observe Figure 6.4.4, which shows the rate of new strains emerging under conditioning the epidemic to not die out. In this figure, we make the comparison not over the absolute time-frame of the epidemic, but instead at the rate of lifetimes. Since the unconditioned process will die out, one would expect to see fewer mutations as time increases. However, we wish to see the expected number of mutation events per unit of individual infectious period.

We see in Figure 6.4.4 the cumulative mean of the expected number of mutation events per unit of individual infectious period with or without conditioning. In both, we see a convergence to a steady level. In the unconditioned version, we see the mutation rate converge to around 0.4 events per unit of infectious period, whereas in the conditioned process this mutation rate is closer to one mutation per unit of infectious period. This may be due to the fact that the only processes which will have survived the conditioning are those which have experienced more frequent mutation events. However, under quasi-stationarity, we also expect the number of infectives to stay steady. Since the strains are bounded by the number of infectives, and we should have one infection per unit of infectious period at equilibrium, we cannot have more than one mutation per unit, agreeing with the convergence of the curve in Figure 6.4.4.

Remark 6.4.5. If $K^*(t)$ is the number of emerged strains up to time t, then we can see that $K^*(t)$ jumps up at the points of an inhomogeneous Poisson process with rate $\beta\theta I(t)$. If we consider the Kolmogorov forward equations, we can see that, using $z(t) = \mathbb{E}[K^*(t)]$ and letting $M(t)$ be the event $M(t) = \{\text{mutation in } (t, t+h]\}$ that

$$\begin{aligned} z(t+h) &= (z(t)+1)\mathbb{P}[M(t)] + z(t)(1-\mathbb{P}[M(t)]) + o(h) \\ z(t+h) &= (z(t)+1)\beta\theta h\mathbb{E}[I(t)] + z(t)(1-\beta\theta h\mathbb{E}[I(t)]) + o(h) \\ \Rightarrow z'(t) &= \beta\theta\mathbb{E}[I(t)] \end{aligned}$$

Solving the ODE with the expression for $\mathbb{E}[I(t)]$ in (3.2.2) and initial condition $z(0) = 1$ gives

$$z(t) = \frac{\beta\theta(e^{(\beta-\gamma)t} - 1)}{\beta - \gamma} + 1$$

For the case conditioned on non-extinction, one can generate the differential equation for $z_1(t) = \theta\mathbb{E}[B(t)|I(t) > 0]$, where $B(t)$ is the total number of infections up to time t. Since each infection is independently a mutation with probability θ, the number

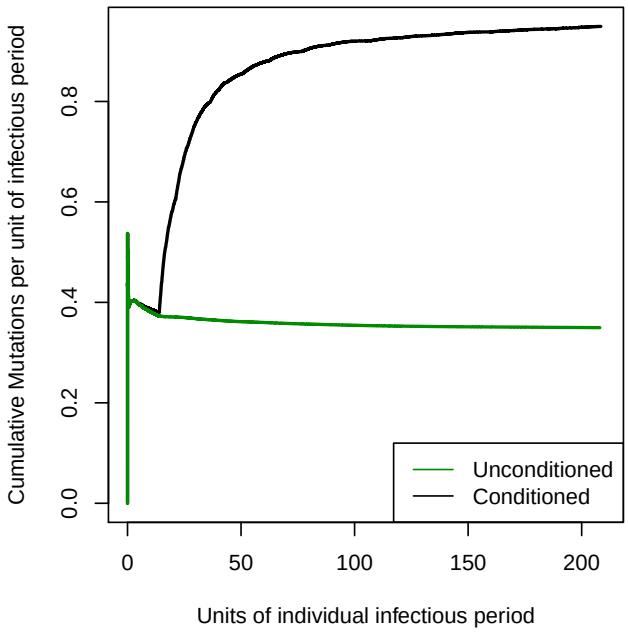


Figure 6.4.4: Number of Emerging Strains Under Conditioning and without Conditioning. $\beta = 2, \gamma = 1, \theta = 0.3, N = 25$. Simulations produced using SMC Sampler: $M = 100$, $T_{\max} = 5$, $T_b = 0$, $T_d = 1$, $T_{\text{end}} = 100$, $\lambda = 0.5$

of emerging strains is precisely $\theta z_1(t)$.

This is done using the Backward Kolmogorov equations and methods extremely similar to Lemmas 4.2.8 and 4.2.7, and the proof of Theorem 4.2.6. Using these we obtain:

$$\begin{aligned} z_1(t+h) &= \mathbb{P}[\text{birth in } (0,h) | I(0)=1, I(t+h)>0]\,(1+\mathbb{E}[B(t)|I(0)=2, I(t)>0]) \\ &\quad + \mathbb{P}[\text{no event in } (0,h) | I(0)=1, I(t+h)>0]\,(1+\mathbb{E}[B(t)|I(0)=1, I(t)>0]) \\ &\quad + o(h) \\ \Rightarrow z_1'(t) &= \left(2\beta(1+z(t)) - \frac{\beta(\beta-\gamma)e^{(\beta-\gamma)t}}{\beta e^{(\beta-\gamma)t}-\gamma}\right) + \frac{\beta(\beta-\gamma)e^{(\beta-\gamma)t}}{\beta e^{(\beta-\gamma)t}-\gamma} z_1(t) \end{aligned}$$

This does not have a closed form solution.

6.5 Limiting Behaviour

One aspect of interest is how the evolving processes relate to those without mutation. To this end, we consider the limits of the times to extinction of the processes as θ tends to 0 or 1, and the limit, for fixed θ, of the time to extinction as N tends to infinity. This would justify using the infinite population models to approximate, for example, the decay parameters for the relevant processes which we cannot obtain analytically.

6.5.1 Limits as Mutation Probability Changes

Theorem 6.5.1. *Let T^θ be the time to extinction of the E-SIS model, and T^1 the time to extinction of the standard SIS model, both starting from a single infective (nominally of strain index 1) in a population of N individuals. Then $T^\theta \to T^1$ in distribution as $\theta \to 1$.*

Proof. To prove this we make use of a coupling of $(T^\theta : 0 < \theta < 1)$ and T^1. We assume the underlying processes to be defined over the same population of individuals indexed by $n = 1, \ldots, N$.

- For each individual n, define a sequence of i.i.d. infectious periods $\{L_m^{(n)}(\omega) \sim \text{Exp}(\gamma) : m \in \mathbb{N}\}$.
- For each *ordered* pair of individuals (n, n'), define a homogeneous Poisson Process $A^{(n,n')}(t, \omega)$ on $[0, \infty)$ with rate β.
- For each ordered pair (n, n') define the indicator dictating whether an infection is a mutation event using a common sequence of uniform random variables $U_l^{(n,n')}(\omega) \sim \text{Unif}[0, 1]$ for $l \in \mathbb{N}$. At the l^{th} contact event, the event is a mutation if and only if $U_l^{(n,n')}(\omega) \leq \theta$.
- Let all $L_m^{(n)}$, $A^{(n,n')}$, $U_l^{(n,n')}$ be independent of each other.

Using these, the coupled associated processes $(\mathbf{Y}, \mathbf{X}^\theta)$ proceed as follows. Starting with an initial infective individual ($n = 1$ without loss of generality), contact events occur at points of the Poisson processes $A^{(1,n')}(t, \omega)$. During individual 1's first infectious period L_1^1, let n' be the first individual that 1 makes contact with. The

SIS model $\mathbf{Y}(t)$ has a successful infection, so individual n' becomes infected. If $U_1^{(1,n')} \geq \theta$, the E-SIS model has a successful infection of individual n' by individual 1 with the same strain index as individual 1 ($k = 1$). If $U_1^{(1,n')} < \theta$, the individual is infected with a new strain $k = 2$. If no such contact takes place during 1's infectious period, then both $\mathbf{X}^\theta$ and $\mathbf{Y}$ die out at the end of the infectious period, and so $T^\theta(\omega) = T^1(\omega)$ for all θ in this case.

During the rest of the epidemic, the models proceed as follows. While a given ordered pair of individuals (n, n') has individual n infected, and n' susceptible, run the corresponding Poisson process $A^{(n,n')}(\omega)$ forwards in time. Pause this process at any time for which these infective statuses are not as above. At the point of infection, the newly infected individual n' stays infected for a period of length $L^{(n')}_{m(n',t)+1}$, where $m(n', t)$ is the number of infections individual n' has recovered from up to the current time t. During an infectious period of individual n, mutation infections occur if $U^{(n,n')}(\omega) < \theta$. In $\mathbf{X}^\theta(t, \omega)$, non-mutation infections of n' by n are only successful if the strain index of n is strictly greater than that of n'. The corresponding infection for $\mathbf{Y}$ is always successful. Mutation infections in $\mathbf{X}$ are always successful, as are the corresponding infections for $\mathbf{Y}$.

Under this coupling we note that, since non-infective contacts do not affect the process, the marginals $\mathbf{Y}(\omega)$ and $\mathbf{X}^\theta(\omega)$ for $\omega \in E$ have the correct distribution by construction.

Fix $\omega \in E$ our probability space. Almost surely, we must have $T^1(\omega) < \infty$. On the interval $[0, T^1)$, there are two possibilities. At each infective-susceptible contact we compare the strain indices. The first option is that every contact can lead to a successful infection, arising from a sequence of infection events which always aim for a susceptible of a lower index. In this case, we have $T^\theta(\omega) = T^1(\omega)$ for all $\theta \in [0, 1]$. The second possibility is that there is at least one such "potentially unsuccessful" contact event. What we mean by this is that, if the event were to be non-mutation, it would fail. This failure occurs if the relevant $U_l^{(n,n')} > \theta$. Since we must have a finite number of such events occurring in $[0, T^1)$, and $U^{(n,n')} \in (0, 1)$, we can find θ_1 such that $U \leq \theta_1$ for all such $U = U_l^{(n,n')}$ corresponding to these potentially unsuccessful events. As mentioned above, this means that for this value of $\theta = \theta_1$ we must have $T^{\theta_1}(\omega) \to T^1(\omega)$ for $\theta \geq \theta_1$. So for every $\omega \in E$, there exists $\theta_1 \in (0, 1)$ such that $T^\theta(\omega) = T^1(\omega)$ for all $\theta \geq \theta_1$. Hence $T^\theta(\omega) \to T^1(\omega)$ as $\theta \to 1$.

This can be done for almost every $\omega \in E$, and hence we obtain the result by the

Skorohod Dudley theorem of Dudley [1968]. □

Theorem 6.5.2. *Let T^0 be the time to extinction of the standard SIR model. Then $T^\theta \to T^0$ in distribution as $\theta \to 0$.*

Intuitively, one can think of identifying the $\mathcal{S}_0$-class for the E-SIS model and the R-class of the SIR model. As mutation events gets rarer, the chance of mutation happening before extinction becomes smaller and smaller, and so the identification of the two processes is more likely to last until extinction. Here we will reuse the calligraphic notation outlined in Section 6.3.1.

Proof. This proof follows in a similar fashion to Theorem 6.5.1. Define a coupling as before of $(T^\theta : 0 < \theta < 1)$ and T^1 via coupling processes $\mathbf{X}^\theta(t,\omega) = (\mathcal{S}(t,\omega), \mathcal{I}(t,\omega))$, the E-SIS model, and $\mathbf{Z}(t,\omega) = (S(t,\omega), I(t,\omega))$ the standard SIR model on a common probability space $(E, \mathcal{F}, \mathbb{P})$ over a common population $\{1,\dots,N\}$. Define the sequences of infection periods $\{L_m^{(n)} : m \in \mathbb{N}\}$, the contact Poisson processes $A^{(n,n')}$, and the mutation indicators $U_l^{(n,n')}$ as in Theorem 6.5.1. In this version, however, the Poisson process $A^{(n,n')}$ progresses during any time for which n is infective, n' is susceptible in $\mathbf{X}$ and n' is susceptible or removed in $\mathbf{Z}$. Infections are only successful if n' is susceptible, and the event is a mutation or n' is of a strictly lower strain index than n. This means that, up to the time of the first secondary infection, if one occurs, the two epidemics are identical if one identifies the $\{\mathcal{S}_1, \mathcal{S}_2, \mathcal{S}_3, \dots\}$ classes for $\mathbf{X}$ and R class for $\mathbf{Z}$.

Fix $\omega \in E$. $T^0(\omega) < \infty$ almost surely, so on the interval $[0, T^0)$, there are two possibilities. Firstly, we consider the case where there are no secondary infections, so no individual in $\{\mathcal{S}_1, \mathcal{S}_2, \dots\}$ in the $\mathbf{X}$ process is contacted and successfully infected. In this case we obtain that $T^\theta(\omega) = T^0(\omega)$ since, under this coupling, whenever there is an infection in $\mathbf{X}$ there is one in Z. More precisely, assuming the initial condition of a single infective of strain 1 and susceptibles of strain index 0, an absence of secondary infections either arises from no mutation events, or happens in the case where mutations happen but the epidemic ends before a relevant infection takes place. The second option is that, for this choice of ω, there is a secondary infection in $\mathbf{X}$ process. At this point, $\mathbf{Z}(t,\omega)$ has one fewer infective than $\mathbf{X}(t,\omega)$. Under the coupling, this means that $T^0(\omega) < T^\theta(\omega)$. The proof follows if, for each ω, we can make a choice of θ such that we are eventually guaranteed to be in the first case.

More precisely, for our choice of ω, there are a finite number of infection contact

events in the interval $[0, T^0)$. For each of these events is a corresponding $U_l^{(n,n')}$ mutation indicator variable for some n, n', l. Due to the finite number of infection contact events, we can choose $\theta_0(\omega)$ such that $\theta_0(\omega) < U_l^{(n,n')}$ for all such U, which means that no infection event (successful or not) is a mutation event. As such, this means $T^{\theta_0}(\omega) = T^0(\omega)$. Since this also holds for all $\theta < \theta_0(\omega)$ we must have $T^\theta(\omega) \to T^0(\omega)$ as $\theta \to 0$. Since this holds for almost every ω we obtain our result almost everywhere again as an application of Skorohod-Dudley of Dudley [1968]. □

We note here that in the case of the E-BDP model, the extinction time is independent of the mutation parameter θ since all infections are successful, and so the number of infectives exactly follows a linear BDP. We make use of this in the following theorem, as this suggests that regardless of θ, there is a unique limit of the extinction time as the population size grows to infinity.

6.5.2 Limits as Population Changes

Here we note that we can make a similar conclusion to one in the standard SIS model, whereby the time to extinction of the SIS model converges to that of the Linear BDP model. Noting that, under a suitable coupling, the time to extinction of the E-BDP has the same extinction time as the Linear BDP without mutation. This leads us to the following result.

Theorem 6.5.3. *Let $T^{\theta,N}$ be the time to extinction of the E-SIS model, and T the time to extinction of the E-BDP model, both starting from a single infective. Then we have $T^{\theta,N} \to T$ in distribution as $N \to \infty$ when $\beta < \gamma$. If $\beta \geq \gamma$, then on the event $\{T < \infty\}$, a region of probability $1 - \gamma/\beta$, we also have $T^{\theta,N} \to T$ in distribution as $N \to \infty$*

Proof. Using Theorems 6.5.1 and 6.5.2 we can conclude for any fixed N that $T^{0,N}$ is the time to extinction for the standard SIR model, and $T^{1,N}$ is equal to the time to extinction for the standard SIS epidemic model. Furthermore, from these theorems we can construct a coupling of the SIS, E-SIS and SIR models using two sets of Poisson processes and mutation indicator variables such that, for any $\theta \in [0, 1]$,

$$T^{0,N}(\omega) \leq T^{\theta,N}(\omega) \leq T^{1,N}(\omega)$$

almost everywhere. This allows us to conclude that the time to extinction also

converges in distribution. From Barbour [1975], we know that, if $\beta < \gamma$ where β is the contact rate and γ the recovery rate, then the time to extinction $T^{0,N}$ converges in distribution to T, the time to extinction of a Linear BDP with the same parameters β, γ. From Andersson and Djehiche [1998] we obtain that the same thing happens for SIS models so that $T^{1,N} \to T$ in distribution as $N \to \infty$. Using the bounding, we obtain therefore that $T^{\theta,N} \to T$ as $N \to \infty$.

In the case where $\beta \geq \gamma$ we note that on a set of probability $1 - \gamma/\beta$, the time to extinction of the linear BDP is infinite, as discussed in Anderson [1991]. From Andersson and Djehiche [1998], we know that $T^{1,N} \to T$ almost surely (and hence in distribution) on the event $\{T < \infty\}$. From Barbour [1975] we know that on this event, $T^{0,N} \to T$ in distribution.

Combining all the above leads to the result. □

It should be noted, that on the event $\{T = \infty\}$ we don't have $T^{0,N} \to \infty$. Instead $T^{0,N}$ converges to an extreme-value distribution as mentioned in Proposition 2.2.4. As such we don't necessarily have $T^{\theta,N} \to \infty$ on $\{T = \infty\}$.

6.6 Reproduction Numbers

To summarise the models defined in the rest of this chapter, we look to a number of key statistics which are related to the commonly used *Basic Reproduction Number*, R_0 which is used to illustrate whether an epidemic is likely to infect a large proportion of the population. To this end, we discuss R_0, a modified version of the *Household Reproduction Number* R_* as defined by Ball [1997], and a modified version of the *Effective Reproduction Number* R_t. These show different aspects of the model, which we will compare and explore within the subsequent simulation study.

6.6.1 Basic Reproduction Number R_0

We first discuss the Basic Reproduction Number as defined in 2.2. This is defined in Anderson and May [1992] as the number of individuals infected by a single typical infective in an otherwise susceptible population. For our model, this equates to a single infective of strain index 0 and $N - 1$ susceptible individuals of strain index -1. Typically, R_0 is calculated for an infinite population, so makes use of branching

process approximations. In the case of the single-strain SIS and SIR models, this equates to computing the expected offspring of a single individual with birth rate β and expected lifetime $1/\gamma$ which gives $R_0 = \beta/\gamma$. In the case of the E-SIS model, we consider the E-BDP. Since the contact rate β does not depend on θ during an infectives lifetime, only the type of the offspring generated, we obtain the same value of $R_0 = \beta/\gamma$, independently of θ.

6.6.2 Modified Household Reproduction Number R_*

One issue with R_0 is that it fails to take into account the likely immunities present in the population, or how much the pathogen evolves during the opening phase of the epidemic. To this end, we use an adjusted version of the household reproduction number as described by Pellis et al. [2012]. In the paper, an epidemic is considered which evolves over a population which is grouped into households, such that individuals in the same household make contact at one rate, and make contact with other households at a potentially different rate. R_* is defined as the expected number of households infected by a single household from an epidemic started from a single individual in that household. This is shown in Pellis et al. [2012] to be equal to

$$R_* = \mathbb{E}[Z]R_H$$

where $\mathbb{E}[Z]$ is the expected number of individuals infected in a single household epidemic, and R_H is the basic reproduction number of an epidemic process where each household in the original epidemic is considered a single individual, contacting other households only. This is approximate for a finite number of households H, but becomes exact in the limit as $H \to \infty$.

In this case, we consider each strain as a "household" which has countably many people in it, and mutations are considered contacts with other such households. In this case $\mathbb{E}[Z]$ is the expected size of a branching process with birth rate $\beta(1-\theta)$ and death rate γ, not including the initial infective. In this case, one can use a geometric series to see that

$$\mathbb{E}[Z] = \begin{cases} \frac{\beta(1-\theta)}{\gamma-\beta(1-\theta)} & \beta(1-\theta) < \gamma \\ \infty & \beta(1-\theta) > \gamma \end{cases}$$

This means that R_* would be infinite if $\beta(1-\theta) > \gamma$. In the other case, we note that

we expect each infective to infect $\beta\theta/\gamma$ other "households" during their lifetime, so $R_H = \beta\theta/\gamma$. As such we obtain that

$$R_* = \begin{cases} \frac{\beta\theta}{\gamma}\frac{\beta(1-\theta)}{\gamma-\beta(1-\theta)} & \beta(1-\theta) < \gamma \\ \infty & \beta(1-\theta) > \gamma \end{cases}$$

In the above one could consider R_0 to be the "intra-strain" basic reproduction number, and R_H to be the "inter-strain" basic reproduction number. With these we can obtain one of three regimes:

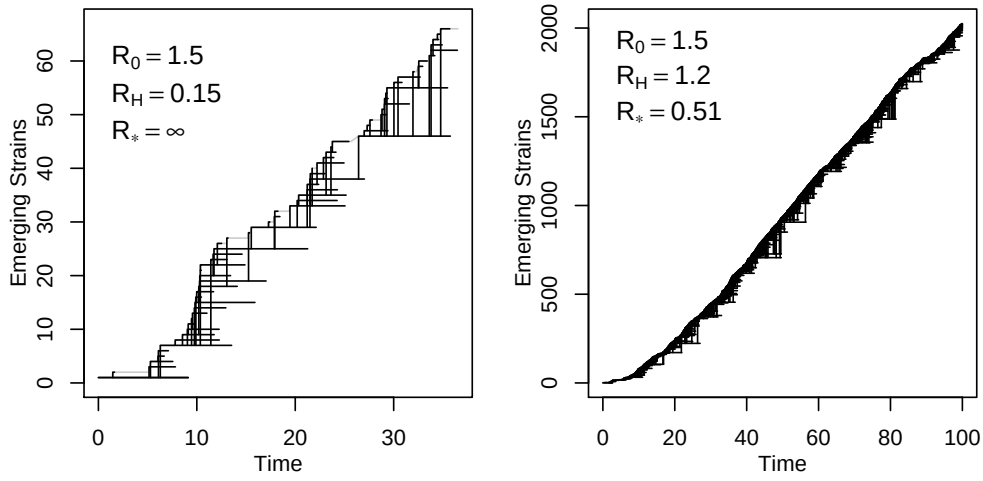
- If $R_0 = \beta/\gamma < 1$, then the whole population would die out with certainty, and no large epidemic would occur.
- If $R_0 \geq 1$ and $R_H < 1$ then the epidemic would, with some probability, reach a large proportion of the population, but each individual strain would die out quickly.
- Thirdly, if $R_0 \geq 1$ and $R_H \geq 1$, then each strain has a positive probability of producing a large outbreak.

Figure 6.6.1 shows the genetic trees under the E-SIS model under the two super-critical regimes, which allows us to see that for small θ, we obtain only a small number of strains, and the epidemic is more likely to die out. Moreover, in a finite population, this low θ leads to high immunity in the population and hence shorter epidemics.

6.6.3 The Effective Reproduction Number

One drawback to the basic reproduction number is that it only usefully describes the initial behaviour of an epidemic when the population is largely susceptible. In the case of endemic conditions where we expect the infection to last a long time, such as seasonal influenza, the proportion of susceptible people drops, and so the chance of a successful infectious contact goes down. One alternative way to describe the infectivity of the epidemic is to consider the *effective reproduction number*, denoted R_t, defined as

$$R_t = R_0 \frac{S(t)}{N}$$



(a) $\beta = 1.5$, $\gamma = 1$, $\theta = 0.1$ (b) $\beta = 1.5$, $\gamma = 1$, $\theta = 0.8$

Figure 6.6.1: Comparisons of Emergence of Strains under different R_0, R_H under the E-SIS model with $N = 25$, $\gamma = 1$, $\beta = 1.5$

in a population of size N. Note that unlike the basic reproduction number, R_t is a random variable dependent on the realisation of the epidemic. Much work has been done in trying to evaluate R_t for specific infections such as influenza by Cowling et al. [2010] and Ebola by Althaus [2014]. However, little has been done on theoretical notions of the distribution of R_t, since in many such epidemic models, calculations for the distribution of $S(t)$ are typically computationally difficult, although some theoretical suggestions have been presented by Nishiura and Chowell [2009] and empirical estimates have been studied by Cintrón-Arias et al. [2009] for specific recorded epidemics.

In the case of a single strain model, the effective reproduction number satisfies the property that if $R_t < 1$ then the epidemic will not infect many more people, and if $R_t > 1$ then, with some probability, the epidemic will infect a large proportion of the population (typically $O(N)$). This can be thought of as restarting the epidemic in the population, but with the current infectious status of the population. At equilibrium, it satisfies $R_t = 1$, which means that the number of infectives is stabilizing but that many more infections may occur.

6.6.4 Quasi-Stationary Reproduction Number R_Q

One problem with R_t is that it is a time-dependent variable which often requires either data or simulations to compute. We offer an alternative based on QSDs which describes the long-term behaviour of the process conditional on non-extinction whilst also being easy to obtain from the QSDs of relevant processes.

We note that, under the evolving strains models, the total number of infectives is always less than the corresponding model without evolving strains. For the E-SIS and E-SIRS models we define the Quasi-Stationary Reproduction Number.

We could make the assumption the typical individual and the otherwise susceptible population are drawn at random assuming infection still exists in the population, so we draw from the quasi-stationary distribution. This mimics to an extent the loss of infections but not of immunity in seasonal influenza. In the lull between 'flu seasons, one might expect only a very small number of infectives. However, the immunity of the population is preserved from the end of one season into the start of the next. This leads us to the following definition.

Definition 6.6.1. The *Quasi-Stationary Reproduction Number* R_Q is defined to be the expected number of secondary infections caused by a single typical individual in an otherwise susceptible population with immunity levels drawn from the QSD.

More precisely, we will draw the single infective individual from the infected marginal of the QSD $\mathbf{u}_I(k)$: the probability that given an individual is infective, it is of strain index k. This is given by

$$\pi_{I,k} = \sum_{\mathbf{x}=(\mathbf{S},\mathbf{I},\mathbf{R})\in\Omega} u_{\mathbf{x}} \frac{I_k}{|\mathbf{I}|}$$

for $k \in \mathbb{K}^* = \{1-N, \dots, 0\}$. Then the susceptibles are drawn from the total strain marginals of the QSD $\mathbf{u}_K(k)$: the probability that under the QSD that a given individual is of strain index k.

$$\pi_{K,k} = \frac{1}{N} \sum_{\mathbf{x}\in\Omega} u_{\mathbf{x}}(I_k + R_k + S_k)$$

This can be thought of as a population which has just become susceptible due to the rare event that all the individuals bar one have recovered and become susceptible, so all strain indices are preserved. Due to the memoryless property of the exponential infectious period, we assume the infected individual is newly infected and at the

start of their infectious period, in a similar fashion to computing R_t.

To arrive at an appropriate value for R_Q, we make use of a similar method used in Section 4.2.2 of Spencer [2007] in modelling immunity and severity. This is similar to the method of Pellis et al. [2012] for obtaining R_0 for multi-strain epidemics which uses eigenvalues of the matrix of expected number of $I_j \to S_k$ infections. Here, for each contacting pair of individuals, we can consider the infectivity of the infective and the susceptibility of the susceptible depending on whether the contact is a mutation or not. We note that, under the equivalence relation described in Section 6.3, we can only have a maximum of N strains in a population of size N, and so the strain index can only range over $k \in \mathbb{K}^* = \{0, 1-N\}$. With this in hand we define the infectivity matrix $A = (a_{ij})_{i=1,2,j\in\mathbb{K}^*}$. Here, $a_{1,j}$ is the expected number of non-mutation contacts made by an individual of strain j during its infectious period, and $a_{2,j}$ is the expected number of mutation contacts made. To compute this, we note that since the probability that a contact is a mutation or not is independent of everything else, and that probability is θ, obtaining that $\theta a_{1,j} = (1-\theta)a_{2,j}$. This means we just need to compute, say, $a_{1,j}$. By independence of contacts, the number of people that a single infective does *not* make contact with during their infectious period is a Binomial random variable $\mathrm{Bin}(N-1, p)$ where p is the probability of not being made contact with. Since we have contact rate between two individuals of β/N and recovery rate of γ, we note that the number of contacts made follow a Poisson distribution $\mathrm{Po}(\beta/\gamma N)$. As such we get that the probability of non-contact is

$$p = \mathbb{P}\left[\mathrm{Po}\left(\frac{\beta}{\gamma N}\right) = 0\right] = e^{\frac{-\beta}{\gamma N}}$$

As such the expected number of non-contacted individuals is

$$(N-1)p = (N-1)e^{\frac{-\beta}{\gamma N}}$$

Using this, the expected number of contacted individuals ρ by an infective of strain k is given by

$$\rho\pi_{I,k} = \pi_{I,k}(N-1)(1-p) \approx \pi_{I,k}\frac{(N-1)\beta}{\gamma N}$$

The second half of the R_Q calculation requires the construction of the susceptibility matrix $B = (b_{j,i})_{j\in\mathbb{K}^*,i=1,2}$. Here $b_{j,1}$ is the probability of a successful infection given a non-mutation contact by an infective of strain j and $b_{j,2}$ is the probability of a successful infection given a mutation contact. Under our model, mutation contacts are always successful and so we immediately obtain $b_{j,2} = 1$ for all j. for $b_{j,1}$ we

make use of the generation of the susceptible population above. Let the probability under the quasi-stationary distribution $\mathbf{u}$ of an individual being susceptible and of strain k be $\pi_{S,k}$. Also let the probability being infective and of strain k is $\pi_{I,k}$, and immune with strain index k by $\pi_{R,k}$. These are given by

$$\pi_{I,k} = \sum_{\mathbf{x}=(\mathbf{S},\mathbf{I},\mathbf{R})\in\Omega} u_{\mathbf{x}} \frac{I_k}{|\mathbf{I}|} \qquad \pi_{R,k} = \sum_{\mathbf{x}=(\mathbf{S},\mathbf{I},\mathbf{R})\in\Omega} u_{\mathbf{x}} \frac{R_k}{|\mathbf{R}|}$$
$$\pi_{S,k} = \sum_{\mathbf{x}=(\mathbf{S},\mathbf{I},\mathbf{R})\in\Omega} u_{\mathbf{x}} \frac{S_k}{|\mathbf{S}|}$$

Then the probability of an individual of our otherwise susceptible population being of strain index k is $\pi_{I,k} + \pi_{S,k}$ in the E-SIS model. In the E-SIRS model the probability is $\pi_{I,k} + \pi_{R,k} + \pi_{S,k}$. Since infections are only successful if the infective is of a greater strain index, then the probability of a successful non-mutation infection is

$$b_{1,j} = \sum_{k=j+1}^{0} \pi_{I,k} + \pi_{S,k} \qquad j \in \mathbb{K}^*$$

Given A and B, the infectivity and susceptibility matrices, we define $M = AB$, an $N \times N$ matrix. which is given by

$$M = AB = \begin{pmatrix} (1-\theta)\rho\pi_{I,(1-N)} & \theta\rho\pi_{I,(1-N)} \\ (1-\theta)\rho\pi_{I,(1-N)+1} & \theta\rho\pi_{I,(1-N)+1} \\ \vdots & \vdots \\ (1-\theta)\rho\pi_{I,0} & \theta\rho\pi_{I,0} \end{pmatrix} \begin{pmatrix} b_{1,(1-N)} & b_{1,-(N-1)+1} & \cdots b_{1,-1} & b_{1,0} \\ 1 & 1 & \cdots 1 & 1 \end{pmatrix}$$

$$= \begin{pmatrix} b_{1,(1-N)}\pi_{I,(1-N)}(1-\theta)\rho + \pi_{I,(1-N)}\theta\rho & \cdots & b_{1,(1-N)}\pi_{I,(1-N)}(1-\theta)\rho + \pi_{I,(1-N)}\theta\rho \\ b_{1,(1-N)+1}\pi_{I,(1-N)+1}(1-\theta)\rho + \pi_{I,(1-N)+1}\theta\rho & \cdots & b_{1,(1-N)+1}(1-\theta)\rho + \pi_{I,(1-N)+1}\theta\rho \\ \vdots & \ddots & \vdots \\ b_{1,0}\pi_{I,0}(1-\theta)\rho + \pi_{I,0}\theta\rho & \cdots & b_{1,0}\pi_{I,0}(1-\theta)\rho + \pi_{I,0}\theta\rho \end{pmatrix}$$

which has the repeated eigenvalue 0 and the single non-zero eigenvalue $\lambda = \rho(\mathbf{b}^T\pi_I(1-\theta) + N\theta)$ where $\mathbf{b} = (b_{1,j})_{j=1-N,\ldots,0}$ and $\pi_I = (\pi_{I,k})_{k=1-N,\ldots,0}$ so

$$R_Q = \rho(\mathbf{b}^T\pi_I(1-\theta) + \theta) \approx \frac{(N-1)\beta}{\gamma N}(\mathbf{b}^T\pi_I(1-\theta) + \theta) \tag{6.6.1}$$

Note that AB and BA have the same non-zero eigenvalues. Here we should note that $\mathbf{b}^T\pi_I$ takes values between 0 and 1. So if $\mathbf{b}^T\pi_I$ is close to 0, then $R_Q \approx \theta\rho$ but if $\mathbf{b}^T\pi_I \approx 1$, then $R_Q \approx \rho$. However, Figure 6.6.2 suggest that $\mathbf{b}^T\pi_I$ is close to zero when θ is small, and so R_Q is small whenever θ is small, since there will be a high

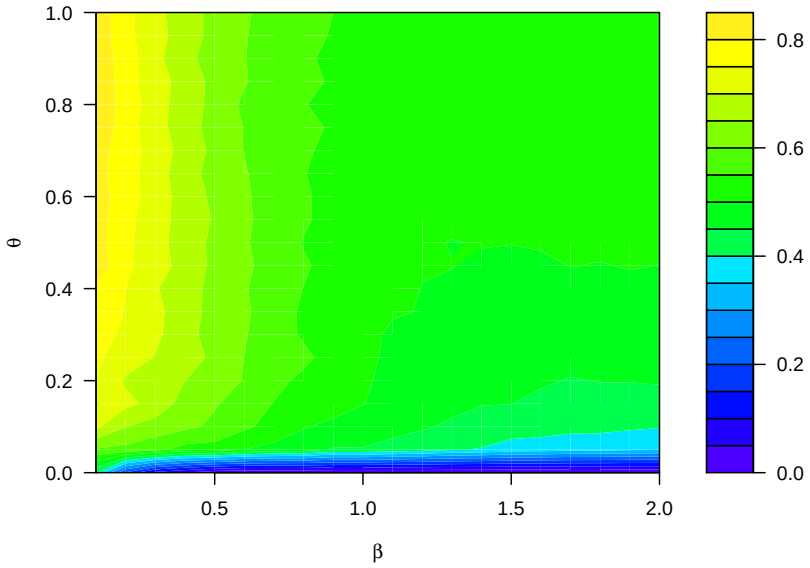
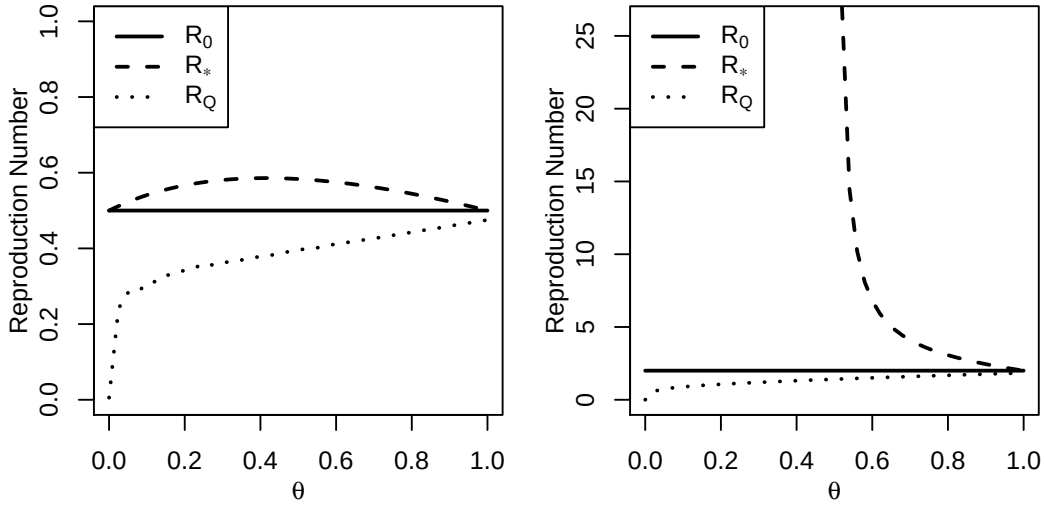


Figure 6.6.2: Illustration of b^* as β, θ change in E-SIS model. Simulations produced using SMC Sampler:
$\gamma = 1$, $N = 25$, $M = 400$, $T_{\max} = 20$, $T_b = 100$, $T_d = 0.25$, $T_{\text{end}} = 140$, $\lambda = 0.4$

amount of immunity in the population due to a low occurrence of mutation events.

As seen in (6.6.1), we note that R_Q increases as β does, and as $\theta \to 1$ we obtain convergence to $\beta\theta/\gamma$. Furthermore, this also happens for $N \to \infty$ as the model behaves more like the E-BDP, or Evolving Transient Immunity process. As θ increases, we also get an increase in R_Q. These all suggest that b^* increases in β and θ, and this is shown in Figure 6.6.2. As δ increases, R_Q does not vary much since we consider the case of starting from a totally susceptible population with a single infective, and so the δ only affects the b^* term in R_Q.

The three notions of a reproduction number in this section describe three different facets of the epidemic model, and can be compared in Figure 6.6.3. It shows that R_Q is always less than R_0 due to the effects of immunity, and R_* depends greatly on θ; the unplotted points for R_* are in the regions where it is infinite, where $\beta(1-\theta) > \gamma$.



(a) $\beta = 0.5$, $\gamma = 1$, $N = 25$ (b) $\beta = 2$, $\gamma = 1$, $N = 25$

Figure 6.6.3: Comparisons of R_0, R_* and R_Q under varying θ.

6.7 Simulation Study

In the final section of this chapter, we use the SMC techniques developed in Chapter 5 to investigate the features of the E-SIS and E-SIRS models which we cannot obtain by standard analytic techniques. We wish to observe how various key statistics perform as we vary the relevant parameters of the model. To this end we look at the following, all of which take expectations over the QSD. This expectation will be denoted by $\mathbb{E}_Q[I_0]$, for example.

- The expected total number of infectives $\mathbb{E}_Q[I]$ and immune individuals $\mathbb{E}_Q[R]$ in the QSD: $I = \sum_{k\in\mathbb{Z}} I_k$, $R = \sum_{k\in\mathbb{Z}} R_k$.
- The expected total number of active strains $\mathbb{E}_Q[K]$ in the QSD where $K = |\{k : I_k > 0\}|$. We also look at how varying the model parameters affects strain diversity in infectives and the whole population.
- The expected proportion of individual-based immunity in the population, ρ_I. This measures the expected proportion of infectives a susceptible is immune to within the population.

$$\rho_I = \sum_{\mathbf{x}=(\mathbf{S},\mathbf{I},\mathbf{R})\in\Omega\setminus\{0\}} u_{\mathbf{x}} \left(\frac{1}{N} \frac{\sum_k (S_k + I_k) \sum_{j\leq k} I_j}{\sum_l I_l} + \frac{\sum_k R_k}{N} \right) \tag{6.7.1}$$

  For example, a susceptible of strain index 1 will have a proportion of individual based immunity of one, since all infectives have a lower strain index. Note this does not take mutation into account. In the E-SIS model the R_k term is omitted as R_k is defined to be zero for all k in this case.

We will focus on the E-SIS model, but also discuss for each statistic how the addition of an immune period, as in the E-SIRS models, changes the number of infectives and strain diversity.

6.7.1 Expected Infectives

We first consider the expected number of infectives in the population under quasi-stationarity, and how this depends on the contact rate, the probability of mutation and the size of the population. We will then discuss how the introduction of an immune period affects this statistic. As discussed and shown in Section 6.4.1, we

note that the introduction of the mutation mechanism, and hence a method of sustained immunity, results in fewer infectives under quasi-stationarity due to the possibility of failed infections. As one might expect, increasing contact rate β or mutation probability θ increases the expected number of infectives, as can be seen from the contour plot in Figure 6.7.1. However, we note that for a fixed population size (in this case $N = 25$), the number of infectives only increases linearly in β when $\mathbb{E}_Q[I]$ is much smaller than N. This can be observed in Figure 6.7.2a, which shows that the number of infectives grows more slowly as β increases for large values of β, especially when θ is small so the probability of failed infections is high.

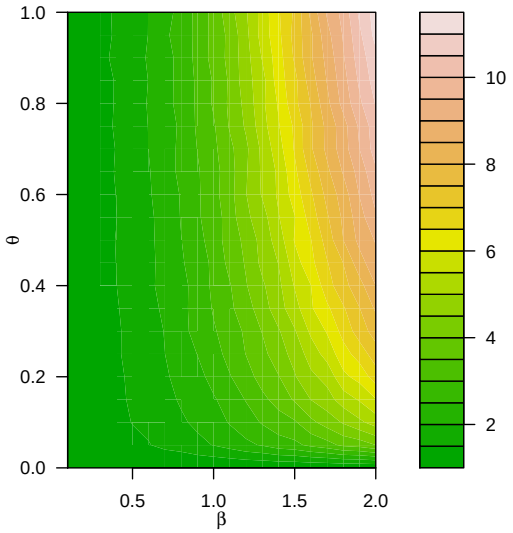
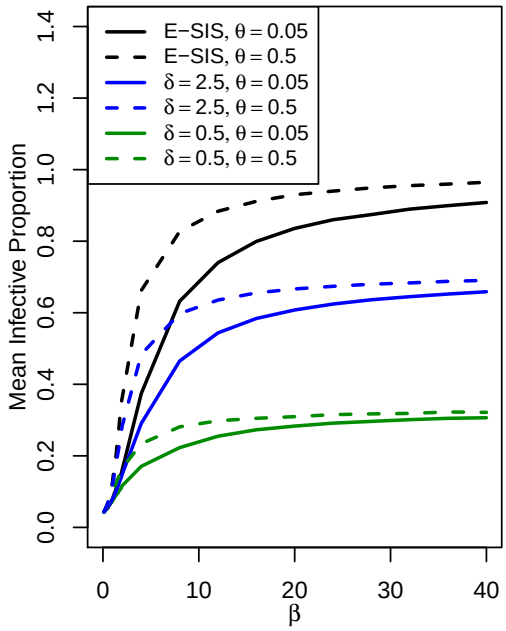


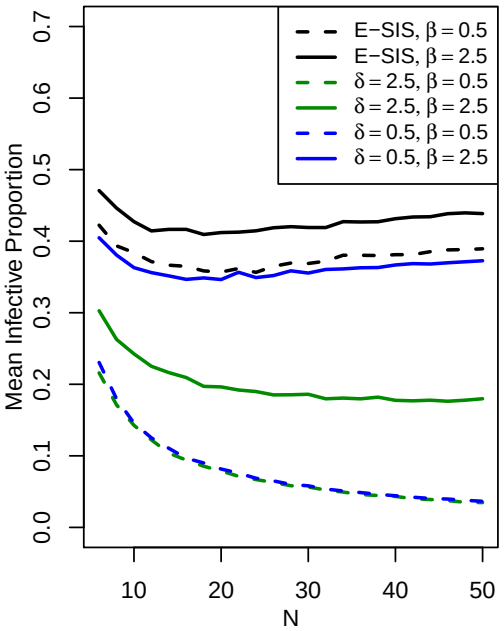
Figure 6.7.1: Mean number of Infectives as β, θ change in E-SIS model. Simulations produced using SMC Sampler:
$\gamma = 1$, $N = 25$, $M = 400$, $T_{\max} = 20$, $T_b = 100$, $T_d = 0.25$, $T_{\text{end}} = 140$, $\lambda = 0.4$

If we consider sending $N \to \infty$, we see in Figure 6.7.2b that the expected proportion of infectives ($\mathbb{E}_Q[I]N^{-1}$) decreases more slowly as $N \to \infty$ in the case where $\beta < \gamma$, whereas in the supercritical case we see that $\mathbb{E}_Q[I]N^{-1}$ keeps increasing, but potentially converges to a second mode in the same fashion as the SIS approximations in Theorem 3.4.2.

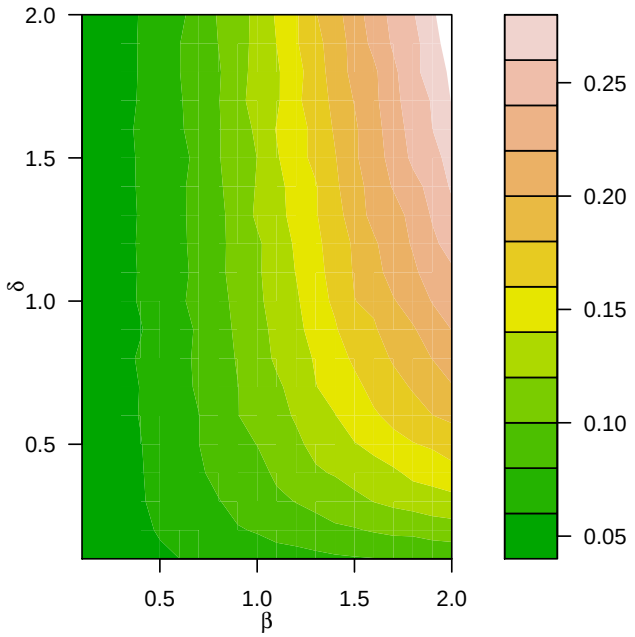
By considering the E-SIRS model, we see that $\mathbb{E}_Q[I]$ is further decreased with the introduction of this transient global immunity. Furthermore, as δ gets smaller we note that since transient immunity lasts longer, $\mathbb{E}_Q[I]$ further decreases. Figure 6.7.2c suggests by the non-symmetric contours that change in δ affects $\mathbb{E}_Q[I]$ less than change in β. If we were to condition instead on $\{S < N\}$, as in Theorem 6.4.2, $\mathbb{E}_Q[R] = 0$ if δ is less than the decay parameter of the E-SIRS model.



(a) Expected infective proportion for fixed $\gamma = 1$, $N = 25$ as β varies.



(b) Expected infective proportion for fixed $\beta = 2$, $\gamma = 1$, $\theta = 0.4$ as N varies



(c) Expected infective proportion as β, δ change in E-SIRS model. $\gamma = 1$, $N = 25$.

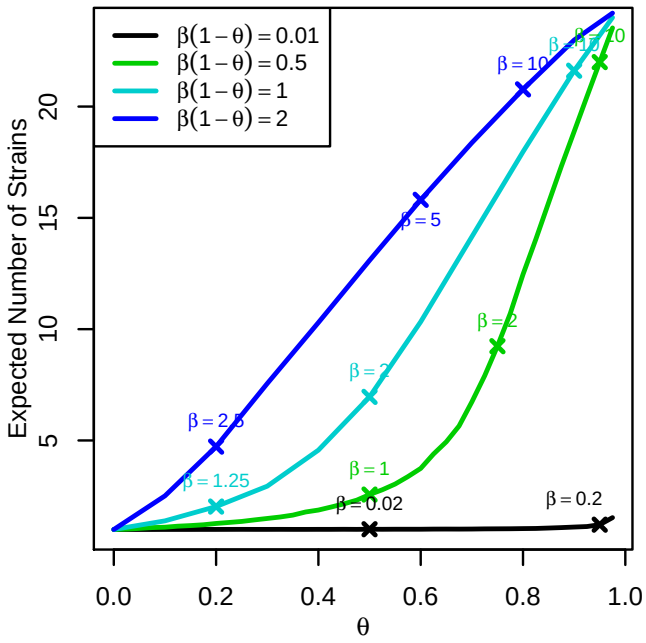
Figure 6.7.2: Expected infective proportion in E-SIS and E-SIRS models for varying values of β, N. Simulations produced using SMC Sampler:
$M = 400$, $T_{\max} = 20$, $T_b = 100$, $T_d = 0.25$, $T_{\text{end}} = 140$, $\lambda = 0.4$

6.7.2 Expected Number of Strains

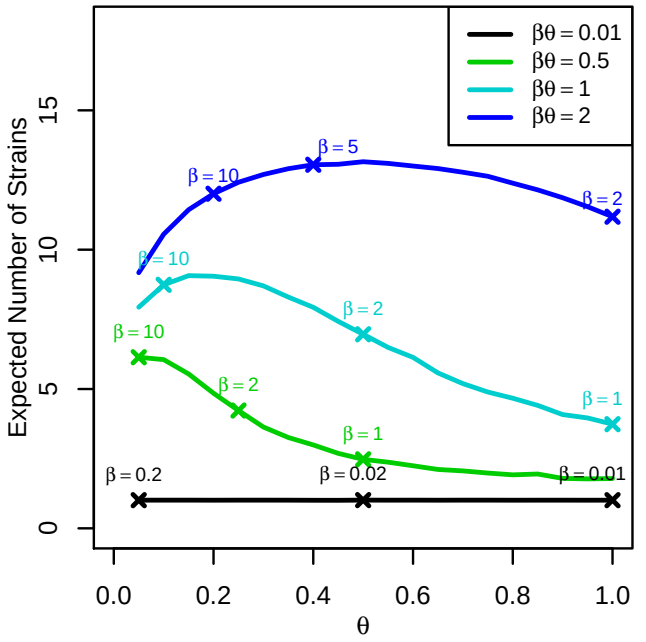
Secondly, we will investigate what happens to the expected number of active strains (strains held by infectives) as the parameters change. We first acknowledge that, under our models, the number of strains is always simultaneously bounded by the number of infectives due to the absence of super-infectivity (infection of an individual by multiple strains during a single infectious period). As such, much of the behaviour is similar to that of the infectives in the previous subsection. Similar to the expected number of infectives, the expected number of strains increases linearly when $\mathbb{E}_Q[K]$ is much less than N. As one would expect, the number of strains increases as the probability of mutation increases whilst $\mathbb{E}_Q[K]$ is much less than N. This follows since we already know that for $\theta = 1$ every infective begins a new strain and as such $\mathbb{E}_Q[I] = \mathbb{E}_Q[K]$. At the other end of the scale, we automatically have that $\mathbb{E}_Q[K] = 1$ if $\theta = 0$.

Again, in the case of the E-SIRS model, it can be shown that the expected number of strains in circulation is decreased by the introduction of a globally immune period. This change is very similar to that for the expected number of infectives, and produces a very similar graph to Figure 6.7.2a.

We investigate in Figure 6.7.3 what happens to the number of strains if we fix the "non-mutation contact" rate $\beta(1-\theta)$, and vary β and θ along this curve. We also consider fixing the "mutation contact" rate $\beta\theta$, and again varying the two parameters. This is to demonstrate two aspects: which of β and θ have a greater effect on the number of circulating strains, and how different values of $\beta(1-\theta)$ affect the number of strains. For example, if one were to consider when $\beta(1-\theta)$ is high, one might expect the number of infectives to be higher, yet the number of strains to be lower. In the case when $\beta\theta$ is high, one might expect $\mathbb{E}_Q[I]$ and $\mathbb{E}_Q[K]$ to be closer in value. This is demonstrated in Figure 6.7.3b, where we see that for fixed $\beta\theta$, the number of strains and infectives are much closer for the $\beta\theta = 2$ line than for the $\beta\theta = 0.05$ line. We also note that there exists a maximum point for the number of strains as β increases after which the number of strains decreases and plateaus. In Figure 6.7.3a, we see that as $\beta(1-\theta)$ increases, the number of strains becomes more linear in θ, and this is very similar for the number of infectives. Note that for this figure, both θ and β increase from left to right, whereas, to maintain fixed $\beta\theta$, θ decreases as β increases. We shall return to these fixed "mutation contact" and "non-mutation" contact rates when discussing individual immunity.



(a) Expected Number of Strains for fixed $\beta(1-\theta)$, $\gamma = 1$, $N = 25$.



(b) Expected Number of Strains for fixed $\beta\theta$, $\gamma = 1$, $N = 25$.

Figure 6.7.3: Expected Number of Strains for fixed mutation contact rates $\beta\theta$ in the E-SIS and E-SIRS models. Simulations produced using SMC Sampler: $M = 400$, $T_{\max} = 20$, $T_b = 100$, $T_d = 0.25$, $T_{\text{end}} = 140$, $\lambda = 0.4$

6.7.3 Strain Diversity

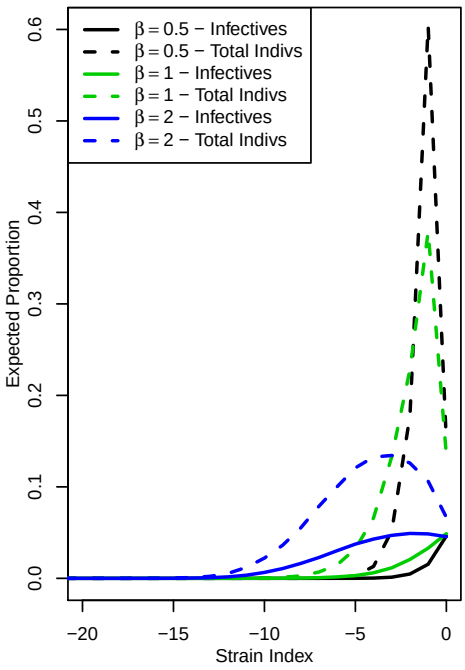
In a similar fashion to Section 6.7.2, we still focus on the number of strains. However, we now consider the number of individuals with a given strain under quasi-stationarity. We see in Figure 6.7.4 what happens as β, θ and N change. In Figure 6.7.4a, we see that as β increases there are more infectives, and a more pronounced curve in the strain diversity. Here we have chosen a high value for θ to more clearly demonstrate this. Another point of interest is the lag of the strain diversity: the number of strains between the peak infective strain (the mode of the infective strains) and the peak of the immunity (the mode of the susceptible strains). We see that the lag is fairly consistent for the different values of β, but does increase slowly for large β.

If we turn our attention to Figure 6.7.4b, we see that θ has a more profound effect. As θ increases, we see that the number of strains present increases, so the strain diversity curve flattens because $\mathbb{E}_Q[K] \leq \mathbb{E}_Q[I] \leq N$. We also see that possibly because of this, we also observe a larger lag for high values of θ. This may be partly due to the more rapid emergence of new strains.

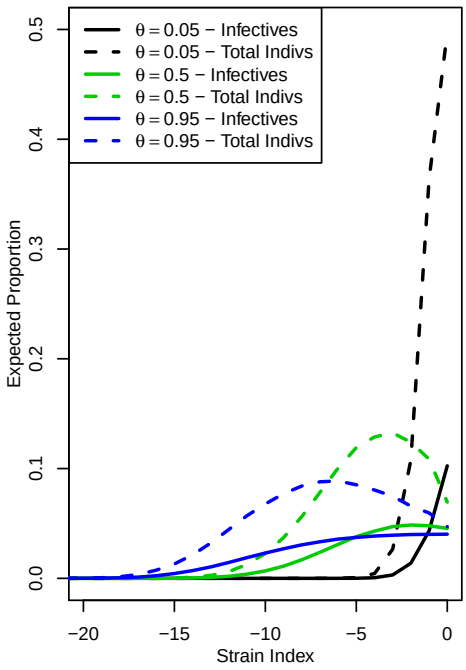
Finally, we see in Figure 6.7.4c that as N increases, we observe a wider number of strains, as one would expect given $\mathbb{E}_Q[K]$'s behaviour. However, unlike the behaviour as β changes, we see that the peak moves away from 0 but the lag appears consistent. Additionally, the mode becomes less pronounced in the susceptible immunity curve as N increases.

We also give some illustration of the strain diversity curve for the E-SIRS model. Note in Figure 6.7.4d that, as in the examples in Section 6.4, the immune period reduces strain diversity by reducing the expected number of infectives.

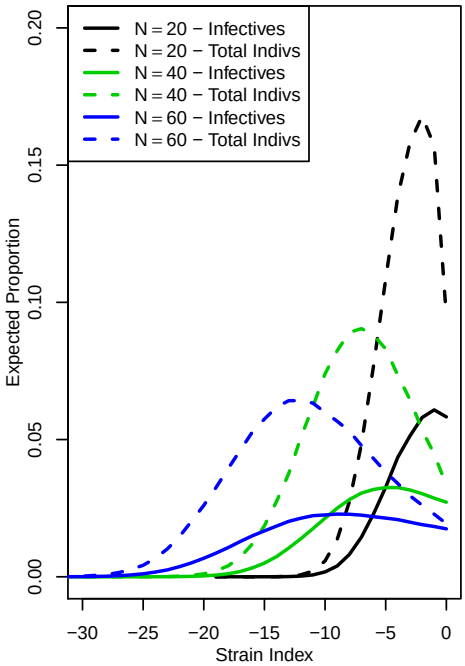
In applications, one might wish to look further into connections between the lag and the model parameters more closely. If one models a pathogen with a large lag, then one may need to vaccinate a large proportion of the population, since the active strain is a long way removed from the current "image" of the pathogen in the immune systems of individuals in the population. Conversely, if the lag is very small, then we may be able to make use of the fact that the currently observed strains in infectives (an easier piece of data to obtain) are more similar to the current "image" of the pathogen in the susceptible population's immunity profile.



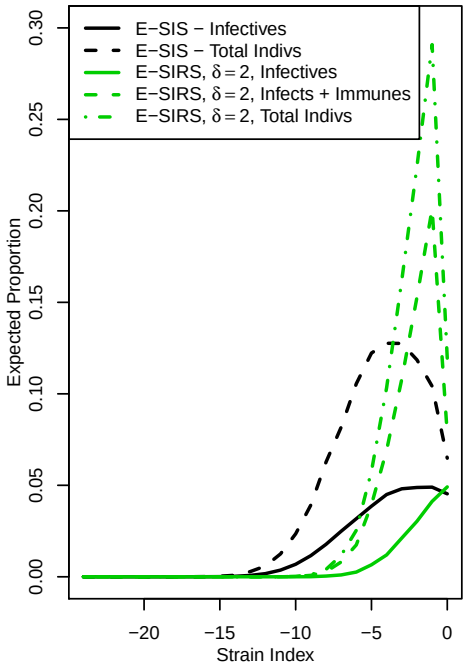
(a) Strain Diversity as β varies. $\theta = 0.4$, $\gamma = 1$, $N = 25$



(b) Strain Diversity as θ varies, $\beta = 2$, $\gamma = 1$, $N = 25$.



(c) Strain Diversity as N varies, $\beta = 2$, $\gamma = 1$, $\theta = 0.5$.



(d) Strain Diversity as β varies, $\beta = 2$, $\gamma = 1$, $\theta = 0.5$, $N = 25$.

Figure 6.7.4: Strain Diversity as key parameters change in the E-SIS and E-SIRS models. Simulations produced using SMC Sampler:
$M = 400$, $T_{\max} = 20$, $T_b = 100$, $T_d = 0.25$, $T_{\text{end}} = 140$, $\lambda = 0.4$

6.7.4 Individual-based Immunity

The next statistic we observe is ρ_I, the level of individual-based immunity within the population. The statistic, as defined in (6.7.1), describes the average proportion of the population to which an individual is immune. It should be noted that this statistic is by construction volatile, and sensitive to Monte Carlo error. This is because an extra mutation event can greatly affect ρ_I since all individuals will become more susceptible. This volatility would be reduced for very large values of N, and for very large particle simulations. Since mutation events affect all infectives and susceptibles equally, the inclusion of this aspect would add on a constant factor dependent on θN^{-1}.

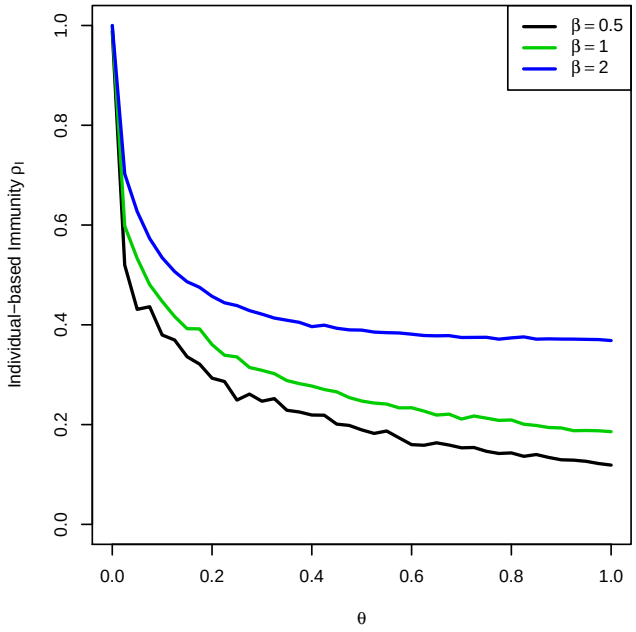
As β increases, ρ_I increases, since there is less time that an individual stays susceptible before being successfully infected again, be that via mutation contact or not. As such, there is less time for the susceptible's strain index to fall too low.

The effects of θ changing are shown in Figure 6.7.5a. We note that at $\theta = 0$ we obtain complete immunity for the population in the E-SIS model, since this corresponds to the SIR model, which has a quasi-stationary distribution with no susceptibles, and hence total population immunity. On the other hand, as $\theta \to 1$, we see that ρ_I decreases, and for large values of β we also note that ρ_I converges to a value above zero. We should also note, that since every individual is immune to their current strain, that $\rho_I \geq 1/N$. Even though we know that for $\theta = 1$ we have a different strain index for every infective, this is not necessarily the case for the susceptibles, as infectives recovering with strain indices in the middle of the strain spectrum will cause more susceptibles to bunch up into a single index.

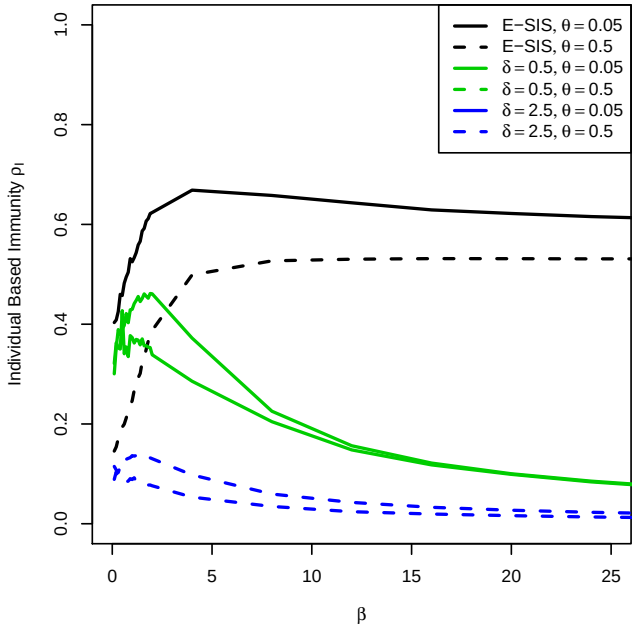
If we include the globally immune period and consider the E-SIRS model, then we see in Figure 6.7.5b that for low values of δ, we experience a much higher level of immunity as one might expect. For larger values of δ we see ρ_I get closer to the curves for the E-SIS model since much less time is spent globally immune.

6.8 Conclusions

In this section we defined a model which can realistically capture aspects of antigenic drift for seasonal strains of Influenza A, despite only depending on just 4 parameters. In fact, if we see γ as just a time-scaling parameter, the model could



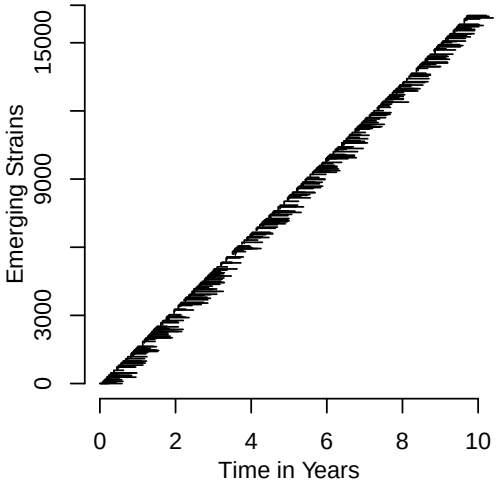
(a) Individual Immunity as θ varies. $\gamma = 1$, $N = 25$.



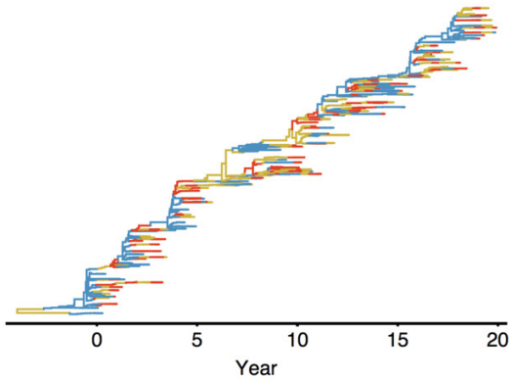
(b) Individual Immunity as δ varies. $\gamma = 1$, $\beta = 2$, $N = 25$.

Figure 6.7.5: Individual based Immunity as δ and θ vary in the E-SIS and E-SIRS models. $\gamma = 1$, $N = 25$. Simulations produced using SMC Sampler: $M = 400$, $T_{\max} = 20$, $T_b = 100$, $T_d = 0.25$, $T_{\text{end}} = 140$, $\lambda = 0.4$

be thought of as only depending on 3 parameters. Compared to models used by Bedford et al. [2015] and Parisi et al. [2013], which require the storage of a whole antigenic history, the model discussed in this chapter is much simpler, which makes simulation, computation and inference much easier. However, despite these simplifications, we see in Figure 6.8.1 the similarity between the simulated genetic tree of H3N2 in Bedford et al. [2015] and the tree generated by the E-SIS model, suggesting that the notion of a new strain which the population is totally susceptible to could be used more in the future. In 6.8.1b, the vertical axis is genetic difference, which was generated by having random jumps at mutation events. It should be noted that the population is much smaller for the demonstration, and the simulation runs over a short time period. This is due to computational limitations with regards to finding an appropriate simulation. However, the overall genetic history of the two pathogens is noticeably similar.



(a) Genetic Tree of E-SIS model, $\beta = 0.88$, $\gamma = 0.2$, $\theta = 0.0004$, $N = 1.5 \times 10^4$, showing top 1% longest lived strains.



(b) Genetic Tree from Extended Figure 7 in Bedford et al. [2015] with permission from publisher.
$\beta = 0.88$, $\gamma = 0.2$, $\theta \approx 0.0004$, $N = 1.5 \times 10^7$

Figure 6.8.1: Comparison of genetic trees from E-SIS with those of Bedford et al. [2015].

The relative simplicity of the model enables analytical insights into model behaviour, such as the relationship between SIS and SIR models discussed in Theorems 6.5.2 and 6.5.1.

Chapter 7

Conclusion

In this thesis, we started with the simplest epidemic models, the SIS and SIR epidemic models (Section 2.2), and a small tool-kit of results regarding quasi-stationary distributions on discrete state spaces (Section 2.1). Using these, we have extended these results to rigorously prove accuracy of approximations (Section 3.4), and characterised the set of QSDs for birth-death and pure death processes (Section 3.3); how the different QSDs relate and how the properties change between them.

We then extended the Linear BDP to define the Transient Immunity model in Chapter 4 which offers the natural extension of the SIRS model to an infinite population. Applying known properties of the Linear BDP and how they relate to the Transient Immunity process allowed the proof of corresponding existence and uniqueness results. However, the addition of a second "dimension" brought up the issue of how the idea of absorption could be interpreted in different ways: having no infectives in the population (Section 4.2) versus having no immunity or infectives in the population (Section 4.3). This was discussed and the two notions were developed and corresponding results on existence and uniqueness were found, as well as results on linking the marginals of the Transient Immunity process to the full QSDs of the Linear BDP.

In this journey, the persistent problem of non-tractability of some key statistics and variables presented itself. To this end, existing SMC methods (defined in Section 2.3) were applied and modified to allow convergence to limiting conditional, rather than stationary, distributions in Chapter 5. More specifically, the key notion of resampling was focused on, and novel techniques were devised and demonstrated.

Particle Refilling (Section 5.2) allowed the continued usage of particles, reducing the workload for the resampling mechanism, and only replacing particles of little or no use: particles corresponding to "absorbed" epidemics. Combine-Split resampling (Section 5.3) reinforced this by using properties of processes on discrete state-spaces to redistribute particles without changing the weight at locations, preserving sample statistics. Regional resampling (Section 5.4) zooms out and looks at maintaining particle weight within regions of the state space, rather than on specific points. To ensure these resampling techniques always perform as expected, Stopping Time Resampling was made use of in Section 5.5 to force resamples whenever any region was lacking in particles.

Chapter 6 took the next step with our models and incorporated the idea of transient immunity into epidemic models with evolving strains to define the E-SIS, E-SIRS, E-BDP and Evolving Transient Immunity models in Section 6.2. Equivalence relations took a process on an uncountable state space with no clear sign of convergence, and allowed one to analyse the current strain diversity and immunity profile of the population. As before, we used links back to the BDP and Transient Immunity models to determine existence and uniqueness results regarding the QSDs of the evolving strain models (Section 6.4). Observing the limits of certain random variables further linked these complicated models to their simpler predecessors (Section 6.5). Finally, we looked at how these models behaved: through analysis of key reproduction numbers R_0, R_* and a similar reproduction number R_Q in Section 6.6 and simulation studies in Section 6.7. Here we also made comparison with existing results and similar work in the field.

Applications and Future Work

The use of marginals and couplings in Chapter 4 was fruitful in the development of results regarding existence and characterisation results of QSDs related to epidemic models. These results may be generalisable to a wider class of processes and the behaviour of their marginals under different notions of quasi-stationarity. Moreover, these results might be extended to processes evolving on continuous state spaces with absorbing boundaries. Characterisation of the Transient Immunity process more generally may prove applicable to related models such as epidemic SIRS models, and in provoking thought more generally on QSDs on reducible, countable state spaces.

The Evolving strain models in Chapter 6 could be taken and applied to real-world problems. One could try and incorporate Approximate Bayesian Computation to perform parameter estimation, particularly in estimating rates and probabilities of mutation within Influenza A. Alternatively, one could consider extending the models to include time-dependent (seasonal) contact and mutation rates to observe how this affects the genetic evolution of pathogens which evolve in this way. More could be undertaken. Moreover, discussion on the effect of vaccination regimes on QSDs could be fruitful, with regards to the impact of delays in vaccine manufacture and distribution to efficacy. Finally, time-inhomogeneity is an important part of describing the seasonality of diseases such as influenza, however this can be a challenging concept to discuss in conjunction with quasistationarity. However, the models in Chapter 6 could be modified and analysed with time-inhomogeneous contact and mutation rates.

Finally, the Combine-Split and Regional Resampling methods defined in Sections 5.3 and 5.4 have room to be pushed further, applying similar methods to other aspects of Monte Carlo methods, and to continuous space models in the case of Regional resampling. Currently, there are no general methods which allow simulation of $\mathbf{v}$-LCDs where $\mathbf{v}$ is a distribution with infinite mean. This is pertinent in the case of Linear BDPs where x-invariant QSDs as described in Section 3.2 (for $x < \alpha$) correspond to $\mathbf{v}$-LCDs with initial conditions with infinite mean (sometimes called "high energy" QSDs). This is a good starting point as we already have the difference equations to determine the x-invariant QSDs. It has not yet been discussed for the Transient Immunity process or the Evolving Strain models. To do this, one might consider developing adaptive regional resampling where regions are created as the particles explore the state space.

Appendix A

Proof computations

A.3 Proofs from Chapter 3

Proof of Theorem 3.4.3. We obtain the expression of the stationary distribution by induction on i for any fixed value of N. Throughout we make use of the following recurrence relations for $\mathcal{B}(k,l)$:

$$\begin{aligned}\mathcal{B}(i+1,j) &= \mathcal{B}(i,j) + (i+1)\mathcal{B}(i,j-1) \quad \text{for } j = 1,\dots,i\\ \mathcal{B}(i+1,i+1) &= (i+1)\mathcal{B}(i,i)\\ \mathcal{B}(i+1,0) &= \mathcal{B}(i,0)\end{aligned}$$

To obtain the stationary distributions we solve the equations $\pi^T Q = 0$, which have the general form, for $i = 1,\dots,(N-1)$:

$$-N\rho\pi_0 + \gamma\pi_1 = 0 \tag{A.3.1}$$

$$\left[(N-(i-1))\rho + \frac{\beta(i-1)(N-(i-1))}{N}\right]\pi_{i-1} - \left[i\gamma + (N-i)\rho + \frac{i(N-i)\beta}{N}\right]\pi_i + (i+1)\gamma\pi_{i+1} = 0 \tag{A.3.2}$$

$$\left[\rho + \frac{(N-1)\beta}{N}\right]\pi_{N-1} - N\gamma\pi_N = 0 \tag{A.3.3}$$

Note that the solution (up to normalization) can be obtained without the last equation, and also note that if we compute these in terms of π_0, then the system can be solved sequentially, since only one new variable is introduced in each equation.

For ease of notation, we assume $\pi_0 = 1$, but as a constant multiplying factor, this doesn't affect the computation. From the first equation (A.3.1), we get $\pi_1 = \frac{N\rho}{\gamma}\pi_0$. We rearrange the other equations (A.3.2) and (A.3.3) and multiply out the brackets to obtain the expression:

$$\begin{aligned}
\gamma(i+1)\pi_{i+1} = &-\frac{N!}{(N-i)!(i-1)!}\sum_{j=0}^{i-2}\frac{\rho^{i-j}\beta^j}{\gamma^i N^j}\mathcal{B}(i-2,j) \\
&-\frac{N!}{(N-i)!(i-2)!}\sum_{j=0}^{i-2}\frac{\rho^{(i-1)-j}\beta^{j+1}}{\gamma^i N^{j+1}}\mathcal{B}(i-2,j) \\
&+\frac{N!}{(N-i)!(i-1)!}\sum_{j=0}^{i-1}\frac{\rho^{i-j}\beta^j}{\gamma^{i-1}N^j}\mathcal{B}(i-1,j) \\
&+\frac{N!}{(N-(i+1))!i!}\sum_{j=0}^{i-1}\frac{\rho^{(i+1)-j}\beta^j}{\gamma^i N^j}\mathcal{B}(i-1,j) \\
&+\frac{N!}{(N-(i+1))!(i-1)!}\sum_{j=0}^{i-1}\frac{\rho^{i-j}\beta^{j+1}}{\gamma^i N^{j+1}}\mathcal{B}(i-1,j) \qquad \text{(A.3.4)}
\end{aligned}$$

We note that the first three terms of the right hand side will cancel if we rewrite the second term as

$$-\frac{N!}{(N-i)!(i-1)!}\sum_{j=1}^{i-1}\frac{\rho^{i-j}\beta^j}{\gamma^{i-1}N^j}(i-1)\mathcal{B}(i-2,j-1)$$

Looking at each summand in the first three terms, we note that the exponents for ρ, β, and γ, and the factorial terms match. Looking at the $\mathcal{B}(k,l)$ terms we see that for $j = 0$, which only appears in the first and third terms we have:

$$-\mathcal{B}(i-2,0) + \mathcal{B}(i-1,0) = 0$$

For $j = i-1$, which only appears in the second and third terms we have:

$$-(i-1)\mathcal{B}(i-2,i-2) + \mathcal{B}(i-1,i-1) = 0$$

For the other terms we have

$$-\mathcal{B}(i-2,j) - (i-1)\mathcal{B}(i-1,j-1) + \mathcal{B}(i-1,j) = 0$$

And so these all cancel, only leaving us to consider the fourth and fifth terms in (A.3.4). Specifically we currently have:

$$\begin{aligned}\pi_{i+1} &= \frac{N!}{i!(N-(i+1))!}\sum_{j=0}^{i-1}\frac{\rho^{(i+1)-j}\beta^j}{\gamma^i N^j}\frac{\mathcal{B}(i-1,j)}{\gamma(i+1)} \\ &\quad + \frac{N!}{(N-(i+1))!(i-1)!}\sum_{j=0}^{i-1}\frac{\rho^{i-j}\beta^{j+1}}{\gamma^i N^{j+1}}\frac{\mathcal{B}(i-1,j)}{\gamma(i+1)} \\ &= \sum_{j=0}^{i-1}\frac{\mathcal{B}(i-1,j)N!}{(N-(i+1))!(i+1)!}\frac{\rho^{(i+1)-j}\beta^j}{\gamma^{i+1}N^j} + \sum_{j=0}^{i-1}\frac{i\mathcal{B}(i-1,j)N!}{(N-(i+1))!(i+1)!}\frac{\rho^{i-j}\beta^{j+1}}{\gamma^{i+1}N^{j+1}} \\ &= \binom{N}{i+1}\sum_{j=0}^{i-1}\frac{\rho^{(i+1)-j}\beta^j}{\gamma^{i+1}N^j}\mathcal{B}(i-1,j) + \binom{N}{i+1}\sum_{j=1}^{i}\frac{\rho^{(i+1)-j}\beta^j}{\gamma^{i+1}N^j}(i\mathcal{B}(i-1,j-1))\end{aligned}$$

As before, we compare the summands as j runs. allowing us to combine the terms to give:

$$\pi_{i+1} = \binom{N}{i+1}\sum_{j=0}^{i}\frac{\rho^{(i+1)-j}\beta^j}{\gamma^{i+1}N^j}\mathcal{B}(i,j)$$

which is the required expression for π_{i+1}. The choice for π_0 is got simply by summing the π_i, and dividing by the total mass. □

A.4 Proofs from Chapter 4

Theorem (Full version of 4.1.3)**.** *Let* $\mathbf{Y}(0) = (I(0), R(0)) = (1, 0)$ *and let* $\gamma - \beta \neq \delta$ *and* $\gamma \neq \beta$*. Then we have*

$$\begin{aligned}Var(R(t)) = \frac{2\beta\gamma}{(\delta+(\gamma-\beta))(\beta e^{(\beta-\gamma)t}-\gamma)}&\left[\frac{\beta e^{2(\beta-\gamma)t}}{2(\beta-\gamma)} + \frac{\beta e^{((\beta-\gamma)-\delta)t}}{\delta+(\gamma-\beta)}\right. \\ &\left. + \frac{\gamma e^{(\beta-\gamma)t}}{\gamma-\beta} - \frac{\gamma e^{-\delta t}}{\delta}\right] + \\ &\frac{2\beta\gamma}{(\delta+(\gamma-\beta))(\gamma-\beta)}\left[\frac{\beta}{2(\beta-\gamma)} + \frac{\beta}{\delta+(\gamma-\beta)} + \frac{\gamma}{\gamma-\beta} - \frac{\gamma}{\delta}\right]\end{aligned}$$

If $\gamma - \beta = \delta$ *then*

$$Var(R(t)) = \frac{4\beta\gamma}{\delta^3}e^{-\delta t} + \gamma t e^{-\delta t} - \gamma^2 t^2 e^{-2\delta t} - \frac{2\beta\gamma}{\delta^3}\left(\delta^2 t^2 + 2\delta t + 2\right)$$

Finally, if $\gamma = \beta$ then

$$Var(R(t)) = \frac{\gamma^3}{\delta^4}\left(2\delta^2 t + 4\delta(e^{-\delta t} - 1) + (1 - e^{-2\delta t})\right) + \frac{\gamma}{\delta}(1 - e^{-\delta t}) - \frac{\gamma^2}{\delta^2}\left(1 + e^{-2\delta t} - 2e^{-\delta t}\right)$$

Proof of Theorem 4.1.3. We consider the first case where $\gamma - \beta \neq \delta$. Conditioning on the first event, we use the Kolmogorov Forward Equation to see that, letting $g(t) = \mathbb{E}[R(t)^2]$, we get

$$g'(t) = \mathbb{E}_{(2,0)}[R(t)^2]\beta + \mathbb{E}_{(0,1)}[R(t)^2]\gamma - (\beta + \gamma)\mathbb{E}[R(t)^2]$$

Using the branching property, where $\mathbf{Y}(t) = \mathbf{Y}^{(1)}(t) + \mathbf{Y}^{(2)}(t)$ is the sum of two independent processes, we can conclude that

$$\mathbb{E}_{(2,0)}[R(t)^2] = 2\mathbb{E}[R(t)^2] + 2\mathbb{E}[R(t)]^2$$

Secondly, if $\mathbf{Y}(0) = (0, 1)$ then $R(t)$ (and so $R(t)^2$) is the life status of a single exponential random variable so

$$\mathbb{E}_{(0,1)}[R(t)^2] = \mathbb{P}[\text{Exp}(\delta) > t] = e^{-\delta t}$$

Using the above two equations we can rearrange to obtain the ODE

$$g'(t) = (\beta - \gamma)g(t) + 2\beta\mathbb{E}[R(t)]^2 + \gamma e^{-\delta t} \tag{A.4.1}$$

Each of the cases in the statement of the theorem follows by substituting in the relevant expression for $\mathbb{E}[R(t)]$ from Theorem 4.1.1 into (A.4.1) and solving the resultant equation. □

Proof of 4.2.6. Once again we use the more concise notation $\mathbf{Y}(t) = (I(t), R(t))$, and assume that $\mathbf{Y}(0) = (1, 0)$ unless otherwise stated. In addition we introduce the notation

$$y(s) = \mathbb{E}[R(s)|I(s) > 0]$$

We will construct the backwards equation for this process by conditioning on the first event of the process. Since we condition on surviving up to time $t + h$, we know that the first event cannot be a recovery since this would give $I(h) = 0$, which

contradicts the non-extinction.

$$y(t+h) = \mathbb{E}[R(t+h)|I(t+h)>0, \mathbf{Y}(h)=(2,0)]\mathbb{P}[\text{infect in } (0,h)|I(t+h)>0] + \mathbb{E}[R(t+h)|I(t+h)>0, \mathbf{Y}(h)=(1,0)]\mathbb{P}[\text{no event in } (0,h)|I(t+h)>0] + o(h) \tag{A.4.2}$$

Using Bayes' rule we can decompose both $\mathbb{P}[\text{infect in } (0,h)|I(t+h)>0]$ and $\mathbb{P}[\text{no event in } (0,h)|I(t+h)>0]$ into terms we already know the expression for:

$$\begin{aligned}\mathbb{P}[\text{infect in } (0,h)|I(t+h)>0] &= \mathbb{P}[I(t+h)>0|\text{infect in } (0,h)]\frac{\mathbb{P}[\text{infect in } (0,h)]}{\mathbb{P}[I(t+h)>0]} \\ &= \mathbb{P}_{(2,0)}[I(t)>0]\frac{\beta h}{\mathbb{P}[I(t+h)>0]}\end{aligned} \tag{A.4.3}$$

$$\mathbb{P}[\text{no event in } (0,h)|I(t+h)>0] = \mathbb{P}[I(t)>0]\frac{1-(\beta h+\gamma h)}{\mathbb{P}[I(t+h)>0]} \tag{A.4.4}$$

Then substituting all this information back into (A.4.2), and using Lemmas 4.2.7 and 4.2.8 we obtain

$$\begin{aligned}y(t+h) = y(t)P_A(t)(2-P_A(t))\frac{\beta h}{P_A(t+h)} &+ \mathbb{E}[R(t)]\frac{2P_A(t)\beta h}{P_A(t+h)} \\ &+ y(t)\frac{P_A(t)(1-(\beta h+\gamma h))}{P_A(t+h)} + o(h)\end{aligned} \tag{A.4.5}$$

and we construct the differential equation as follows:

$$\begin{aligned}\frac{y(t+h)-y(t)}{h} &= y(t)\left(\frac{P_A(t)}{P_A(t+h)}\left[\frac{1}{h}+(\beta-\gamma)\right] - \frac{2\beta P_A(t)^2}{P_A(t+h)} - \frac{1}{h}\right) \\ &\quad + \mathbb{E}[R(t)]\frac{2\beta P_A(t)}{P_A(t+h)} + \frac{o(h)}{h} \\ &= y(t)\left[\frac{P_A(t)-P_A(t+h)}{hP_A(t+h)} + \frac{P_A(t)}{P_A(t+h)}((\beta-\gamma)-2\beta P_A(t))\right] \\ &\quad + \mathbb{E}[R(t)]2\beta\frac{P_A(t)}{P_A(t+h)} + \frac{o(h)}{h}\end{aligned} \tag{A.4.6}$$

By using continuity of $P_A(t)$ and noting that

$$\lim_{h\to 0}\left(\frac{P_A(t)-P_A(t+h)}{hP_A(t+h)}\right) = \frac{-P_A'(t)}{P_A(t)}$$

we take limits on both sides of (A.4.6) as we let $h \to 0$, and obtain the differential

equation

$$\frac{d}{dt}y(t) = y(t)\left[\frac{-P_A'(t)}{P_A(t)} + (\beta - \gamma) - 2\beta P_A(t)\right] + 2\beta\mathbb{E}[R(t)]$$

We can substitute in the expression for $P_A(t)$ to get

$$\frac{-P_A'(t)}{P_A(t)} = \frac{\gamma(\beta-\gamma)^2 e(\beta-\gamma)t}{(\beta e^{(\beta-\gamma)t})^2}\frac{\beta e^{(\beta-\gamma)t} - \gamma}{(\beta-\gamma)e^{(\beta-\gamma)t}} = \frac{\gamma(\beta-\gamma)}{\beta e^{(\beta-\gamma)t} - \gamma}$$

which, after rearranging, and using the initial conditions $\mathbf{Y}(0) = (1, 0)$ gives the differential equation

$$y'(t) = y(t)\left[\frac{\beta(\beta-\gamma)e^{(\beta-\gamma)t}}{\beta e^{(\beta-\gamma)t} - \gamma}\right] + \frac{2\beta\gamma}{\delta + (\beta-\gamma)}(e^{(\beta-\gamma)t} - e^{-\delta t})$$

This can be solved using an integrating factor to give the general solution

$$\begin{aligned} y(t) = \frac{2\beta\gamma}{(\delta + (\beta-\gamma))(\beta e^{(\beta-\gamma)t} - \gamma)}\Bigg[&\frac{\beta}{2(\beta-\gamma)}e^{2(\beta-\gamma)t} - \frac{\beta}{(\beta-\gamma)-\delta}e^{((\beta-\gamma)-\delta)t} \\ &- \frac{\gamma}{\beta-\gamma}e^{(\beta-\gamma)t} - \frac{\gamma}{\delta}e^{-\delta t}\Bigg] + C \end{aligned}$$

Using the initial condition $y(0) = 0$ we can solve to get

$$C = -\frac{2\beta\gamma}{(\delta + (\beta-\gamma))(\beta-\gamma)}\left[\frac{\beta}{2(\beta-\gamma)} - \frac{\beta}{(\beta-\gamma)-\delta} - \frac{\gamma}{\beta-\gamma} - \frac{\gamma}{\delta}\right]$$

giving the result. □

Proof of Theorem 4.2.9. We do this in the same sort of manner to the arguments seen before using the backward equation. First note that

$$\begin{aligned} \mathrm{Var}(I(t)|I(t) > 0) &= \mathbb{E}[(I(t))^2|I(t) > 0] - \mathbb{E}[I(t)|I(t) > 0]^2 \\ &= \frac{\mathbb{E}[I(t)^2]}{\mathbb{P}[I(t) > 0]} - \frac{\mathbb{E}[I(t)]^2}{\mathbb{P}[I(t) > 0]^2} \end{aligned}$$

We have expressions for $\mathbb{E}[I(t)]$ and $\mathbb{P}[I(t) > 0]$, which leaves us to find the expression for $\mathbb{E}[I(t)^2]$, using the backward equation.

$$\begin{aligned} \mathbb{E}[I(t+h)^2] &= \mathbb{E}[I(t+h)^2|I(h) = 2]\beta h + \mathbb{E}[I(t+h)^2|I(h) = 1](1 - (\beta h + \gamma h)) + o(h) \\ &= \mathbb{E}[(I^{(1)}(t) + I^{(2)}(t))^2|I^{(n)}(h) = 1]\beta h \\ &\quad + \mathbb{E}[I(t+h)^2|I(h) = 1](1 - (\beta h + \gamma h)) + o(h) \end{aligned}$$

Here we split $I(t)$ into $I(t) = I^{(1)}(t) + I^{(2)}(t)$ using the branching property, so the numbered processes are independent and have the same distribution as $I(t)$. Using independence we get

$$\begin{aligned}\mathbb{E}[I(t+h)^2] &= 2\mathbb{E}[I(t)^2]\beta h + 2\mathbb{E}[I(t)]^2\beta h + \mathbb{E}[I(t)^2](1-\beta h-\gamma h) + o(h)\\ &= \mathbb{E}[I(t)^2](1+(\beta h-\gamma h)) + 2\mathbb{E}[I(t)]^2\beta h + o(h)\end{aligned}$$

Rearranging, dividing by h, and taking the limit as $h \to 0$ gives

$$\frac{d}{dt}\mathbb{E}[I(t)^2] = (\beta-\gamma)\mathbb{E}[I(t)^2] + 2\beta e^{2(\beta-\gamma)t}$$

which, with the initial condition $\mathbb{E}[I(0)^2] = 1$ we get

$$\mathbb{E}[I(t)^2] = \frac{2\beta e^{2(\beta-\gamma)t} - (\beta+\gamma)e^{(\beta-\gamma)t}}{\beta-\gamma}$$

Substituting this back in gives

$$\mathrm{Var}(I(t)|I(t)>0) = \frac{(2\beta e^{(\beta-\gamma)t} - (\beta+\gamma))(\beta e^{(\beta-\gamma)t} - \gamma)}{(\beta-\gamma)^2} - \frac{(\beta e^{(\beta-\gamma)t} - \gamma)^2}{(\beta-\gamma)^2}$$

and proves the result. □

Proof of 4.2.11. Using the branching property of the Transient Immunity Process, we decompose $\mathbf{Y}$ into $\mathbf{Y}^{(1)} + \mathbf{Y}^{(2)}$ with absorption times $T^{(n)}$ for $n = 1, 2$.

$$\mathbb{E}_{(2,0)}[R(t)^2|I(t)>0] = \mathbb{E}[(R^{(1)}(t) + R^{(2)}(t))^2|I(t)>0] =: (B)$$

Splitting $\{T > t\}$ as in Lemma 4.2.7 gives

$$\begin{aligned}(B) = \mathbb{E}[(R^{(1)}+R^{(2)})^2|T^{(1)}>t, T^{(2)}\le t]\frac{1-P_A(t)}{2-P_A(t)}\\ + \mathbb{E}[(R^{(1)}+R^{(2)})^2|T^{(1)}\le t, T^{(2)}>t]\frac{1-P_A(t)}{2-P_A(t)}\\ + \mathbb{E}[(R^{(1)}+R^{(2)})^2|T^{(1)}>t, T^{(2)}>t]\frac{P_A(t)}{2-P_A(t)}\end{aligned}$$

Expanding the brackets and using the iid property of the two processes allows us to

rewrite this in terms of the original process

$$(B) = \left(2\mathbb{E}[R(t)^2|T > t] + 4\mathbb{E}[R(t)|T > t]\mathbb{E}[R(t)|T \leq t] + 2\mathbb{E}[R(t)^2|T \leq t]\right)\frac{1 - P_A(t)}{2 - P_A(t)}$$
$$+ \left(2\mathbb{E}[R(t)^2|T > t] + 2\mathbb{E}[R(t)|T > t]^2\right)\frac{P_A(t)}{2 - P_A(t)}$$

As in Lemma 4.2.7 we note that

$$\mathbb{E}[R(t)^2|T \leq t] = \frac{\mathbb{E}[R(t)^2]}{1 - P_A(t)} - \frac{\mathbb{E}[R(t)^2|T > t]P_A(t)}{1 - P_A(t)}$$

Substituting this into (B) obtains the result. □

Proof of Thm 4.2.10. This follows in a similar fashion to Theorems 4.2.14 and 4.1.3. Initial conditions are assumed to be $\mathbf{Y}(0) = (1, 0)$ unless otherwise stated. We look to find the differential equation describing the second moment $\mathbb{E}[R(t)^2|I(t) > 0]$. To this end, we first condition on the first event. Note that, starting from $\mathbf{Y}(0) = (1, 0)$, we need only consider the infection and "no event" cases, since recovery, which gives us $I(h) = 0$, cannot happen under the event $I\{(t + h) > 0\}$.

Let $z(t) = \mathbb{E}[R(t)^2|I(t) > 0]$. Then we obtain:

$$z(t + h) = \mathbb{E}[R(t + h)^2|I(t + h) > 0, \mathbf{Y}(h) = (2, 0)]\mathbb{P}[\text{ infect in } (0, h)|I(t + h) > 0]$$
$$+ \mathbb{E}[R(t + h)^2|I(t + h) > 0]\mathbb{P}[\text{ no event in } (0, h)|I(t + h) > 0] + o(h)$$

Using equations (A.4.3) and (A.4.4), we get

$$z(t + h) = \mathbb{E}_{(2,0)}[R(t)^2|I(t) > 0]\frac{\beta h P_A(t)(2 - P_A(t))}{P_A(t + h)}$$
$$+ \mathbb{E}[R(t)^2|I(t) > 0]\frac{P_A(t)(1 - (\beta h - \gamma h))}{P_A(t + h)} + o(h)$$

Using Lemma 4.2.11, we substitute in $\mathbb{E}_{(2,0)}[R(t)^2|I(t) > 0]$ and rearrange to obtain

$$z(t + h) = z(t)\frac{P_A(t)}{P_A(t + h)}(1 + (\beta - \gamma)h - 4P_A(t)\beta h)$$
$$+ 4\beta h\frac{P_A(t)}{P_A(t + h)}\mathbb{E}[R(t)]\mathbb{E}[R(t)|I(t) > 0] - 2\beta h\mathbb{E}[R(t)|I(t) > 0]^2\frac{P_A(t)^2}{P_A(t + h)}$$
$$+ 2\beta h\mathbb{E}[R(t)^2]\frac{P_A(t)}{P_A(t + h)} + o(h)$$

Dividing by h and again referring to Theorem 4.2.6 which notes that

$$\lim_{h\to 0} \frac{P_A(t)}{hP_A(t+h)} - \frac{1}{h} = \frac{-P_A(t)'}{P_A(t)}$$

we take limits in h to get the result in the statement of the theorem. □

A.5 Proofs from Chapter 5

Proof of Lemma 5.4.5. In what follows we will use the notation

$$u^+ = u + \frac{1-u}{N_2} \qquad\qquad u^- = \frac{(N_1-1)u}{N_1-u}.$$

Here u^+ corresponds to the location of $X(n)$ after jumping up from u as in the second case of (5.4.1). Similarly, u^- corresponds to the location of X after jumping down from u as in the first case of (5.4.1).

To show condition 1, we choose $V(x) = \frac{x}{1-x}$ to be our Lyapunov function. With this function, we have the following.

$$\begin{aligned}
V(x^+) &= \frac{x + \frac{1-x}{N_2}}{1 - (x + \frac{1-x}{N_2})} \\
&= \frac{(N_2-1)x + 1}{(N_2-1)(1-x)} \\
&= V(x) + \frac{1}{(N_2-1)(1-x)} \\
&= \left(1 - \frac{1}{N_2-1}\right) V(x) + \frac{1}{N_2-1} \\
V(x^-) &= \frac{(N_1-1)x}{N_1-x} \frac{1}{1 - \frac{(N_1-1)x}{N_1-x}} \\
&= \frac{(N_1-1)x}{N_1(1-x)} \\
&= \frac{N_1-1}{N_1} V(x)
\end{aligned}$$

Substituting into $PV(x)$ gives:

$$\begin{aligned} PV(x) &= \frac{N_1}{N_1+\delta N_2}\left(\frac{N_1-1}{N_1}V(x)\right) + \frac{\delta N_2}{N_1+\delta N_2}\left(\frac{N_2}{N_2-1}V(x)+\frac{1}{N_2-1}\right) \\ &= \frac{(N_1-1)}{N_1+\delta N_2}V(x) + \frac{\delta N_2^2}{(N_1+\delta N_2)(N_2-1)}V(x) + \frac{\delta N_2}{(N_2-1)(N_1+\delta N_2)} \\ &= \left(1-\frac{N_2(1-\delta)-1}{(N_1+\delta N_2)(N_2-1)}\right)V(x) + \frac{\delta N_2}{(N_2-1)(N_1+\delta N_2)}. \end{aligned}$$

Therefore we set

$$\begin{aligned} \lambda &= 1-\frac{N_2(1-\delta)-1}{(N_2-1)(N_1+\delta N_2)} \\ K &= \frac{\delta N_2}{(N_2-1)(N_1+\delta N_2)} \end{aligned}$$

Since we have the condition that $N_2 > (1-\delta)^{-1}$ it follows that $\lambda < 1$ which satisfies condition 1 of Theorem 5.4.4. □

Proof of Lemma 5.4.6. As in the previous proof we use the notation u^+, u^-:

$$u^+ = u + \frac{1-u}{N_2} \qquad\qquad u^- = \frac{(N_1-1)u}{N_1-u}.$$

Now we define the distance-like function $d(x,y) := \min\left\{\bar{d}(x,y), 1\right\}$ where

$$\bar{d}(x,y) := \left|\frac{1}{1-x}-\frac{1}{1-y}\right|$$

To prove that P is d-contracting, we need only consider x, y such that $d(x,y) < 1$. We proceed by showing $\bar{d}(x^+,y^+) = \frac{N_2}{N_2-1}\bar{d}(x,y)$ and $\bar{d}(x^-,y^-) = \frac{N_1-1}{N_1}\bar{d}(x,y)$. Firstly,

$$\begin{aligned} \bar{d}(x,y) &= \left|\frac{1}{1-x}-\frac{1}{1-y}\right| \\ &= \frac{|x-y|}{(1-x)(1-y)}. \end{aligned}$$

Using this we see

$$\begin{aligned}
\bar{d}(x^+, y^+) &= \left| \frac{1}{1 - x - \frac{(1-x)}{N_2}} - \frac{1}{1 - y - \frac{(1-y)}{N_2}} \right| \\
&= \frac{(1 - \frac{1}{N_2})|x - y|}{\left(1 - x - \frac{(1-x)}{N_2}\right)\left(1 - y - \frac{(1-y)}{N_2}\right)} \\
&= \frac{(1 - \frac{1}{N_2})|x - y|}{(1 - \frac{1}{N_2})^2(1 - x)(1 - y)} \\
&= \frac{N_2}{N_2 - 1}\bar{d}(x, y).
\end{aligned}$$

Additionally, we see

$$\begin{aligned}
\bar{d}(x^-, y^-) &= \left| \frac{1}{1 - \frac{(N_1-1)x}{N_1 - x}} - \frac{1}{1 - \frac{(N_1-1)y}{N_1 - y}} \right| \\
&= \left| \frac{N_1 - x}{N_1 - x - N_1 x + x} - \frac{N_1 - y}{N_1 - y - N_1 y + y} \right| \\
&= \left| \frac{N_1 - x}{N_1(1 - x)} - \frac{N_1 - y}{N_1(1 - y)} \right| \\
&= \frac{1}{N_1} \left| \frac{(N_1 - x)(1 - y) - (N_1 - y)(1 - x)}{(1 - x)(1 - y)} \right| \\
&= \frac{N_1 - 1}{N_1} \frac{|x - y|}{(1 - x)(1 - y)} = \frac{N_1 - 1}{N_1}\bar{d}(x, y).
\end{aligned}$$

To show that P is d-contracting (condition 2 of Theorem 5.4.4), we consider the case where $d(x, y) < 1$. To bound the Wasserstein-d distance needed for d-contraction by we consider the coupling $\widetilde{\pi}$ of two copies of the X-process (X, Y) starting from potentially different starting positions (x, y) and moving the same direction at each step: $(X, Y) \mapsto (X^+, Y^+)$ or $(X, Y) \mapsto (X^-, Y^-)$. This gives

$$\begin{aligned}
\mathcal{W}_d(P(x, \cdot), P(y, \cdot)) &= \inf_{\pi} \int_{E \times E} d(u, v)\pi(du, dv) \\
&\leq \int_{E \times E} d(u, v)\widetilde{\pi}(du, dv) \\
&= \frac{N_1}{N_1 + \delta N_2} d(x^-, y^-) + \frac{\delta N_2}{N_1 + \delta N_2} d(x^+, y^+)
\end{aligned}$$

where for fixed x, y

$$\widetilde{\pi}(du, dv) = \frac{N_1}{N_1 + \delta N_2} \mathbb{1}_{\{u = x^-, v = y^-\}} + \frac{\delta N_2}{N_1 + \delta N_2} \mathbb{1}_{\{u = x^+, v = y^+\}}.$$

However,

$$\begin{aligned}
&\frac{N_1}{N_1+\delta N_2}d(x^-,y^-)+\frac{\delta N_2}{N_1+\delta N_2}d(x^+,y^+)\\
&=\frac{N_1}{N_1+\delta N_2}\bar{d}(x^-,y^-)+\frac{\delta N_2}{N_1+\delta N_2}\min\left\{\bar{d}(x^+,y^+),1\right\}\\
&=\frac{N_1}{N_1+\delta N_2}\frac{N_1-1}{N_1}\bar{d}(x,y)+\frac{\delta N_2}{N_1+\delta N_2}\min\left\{\frac{N_2}{N_2-1}\bar{d}(x,y),1\right\}\\
&\leq\frac{N_1-1}{N_1+\delta N_2}\bar{d}(x,y)+\frac{\delta N_2^2}{(N_1+\delta N_2)(N_2-1)}\bar{d}(x,y)\\
&=d(x,y)-\frac{N_2(1-\delta)-1}{(N_2-1)(N_1+\delta N_2)}d(x,y)\\
&=c\;d(x,y),
\end{aligned}$$

where $c=1-\frac{N_2(1-\delta)-1}{(N_2-1)(N_1+\delta N_2)}$. We have $c<1$ as long as $N_2>\frac{1}{1-\delta}$. Therefore condition 2 of Theorem 5.4.4 is satisfied in this case. □

Proof of Lemma 5.4.7. For condition 3 of Theorem 5.4.4 we must show that $S=\{u:V(u)<4K\}$ is d-small, that is, there exists $s\in(0,1)$ such that for all $x,y\in S$, $\mathcal{W}_d(P(x,\cdot),P(y,\cdot))<s$. For our choice of V and d

$$\begin{aligned}
S&=\{u:V(u)\leq 4K\}\\
&=\left\{u:u\leq\frac{4K}{4K+1}\right\}\\
&=\left\{u:u\leq\tilde{u}:=\frac{4\delta N_2}{4\delta N_2+(N_2-1)(N_1+\delta N_2)}\right\}
\end{aligned}$$

First we show that $d(0,\tilde{u})<1$.

$$\begin{aligned}
d(0,\tilde{u})&=\left|1-\frac{1}{1-\tilde{u}}\right|\\
&=\left|1-\frac{4\delta N_2+(N_2-1)(N_1+\delta N_2)}{(N_2-1)(N_1+\delta N_2)}\right|\\
&=\frac{4\delta N_2}{(N_2-1)(N_1+\delta N_2)}\\
&<\frac{4}{N_2-1}\\
&\leq 1,
\end{aligned}$$

when $N_2 \geq 5$. Then since $d(0, \tilde{u}) < 1$ and using condition 2,

$$\begin{aligned} \mathcal{W}_d(P(x,\cdot), P(y,\cdot)) &\leq c\ d(x,y) \\ &\leq c\ d(0,\tilde{u}) \\ &=: s \in (0,1). \end{aligned}$$

and so S is d-small and condition 3 of Theorem 5.4.4 is satisfied. □

Bibliography

H. Abbey. An examination of the Reed-Frost theory of epidemics. *Human Biology*, 24(3):201, 1952.

C.L. Althaus. Estimating the reproduction number of Ebola virus (EBOV) during the 2014 outbreak in West Africa. *PLoS Currents*, 6, 2014.

R.M. Anderson and R.M. May. *Infectious Diseases of Humans: Dynamics and Control*. Oxford University Press, 1992.

W. Anderson. *Continuous-Time Markov Chains: An Applications-Oriented Approach*. Springer-Verlag, 1991.

H. Andersson and T. Britton. *Stochastic Epidemic Models and Their Statistical Analysis*. Springer, 2000.

H. Andersson and B. Djehiche. A threshold limit theorem for the stochastic logistic epidemic. *Journal of Applied Probability*, 35(3):662–670, 1998.

J.R. Artalejo. On the time to extinction from quasi-stationarity: A unified approach. *Physica A: Statistical Mechanics and its Applications*, 391(19):4483–4486, 2012.

K. Athreya and P. Ney. *Branching Processes*. Springer, 1972.

N.T.J. Bailey. *The Mathematical Theory of Epidemics*. Hafner, 1975.

F. Ball. Epidemics in populations with two levels of mixing. *Annals of Applied Probability*, 7:46–89, 1997.

A.D. Barbour. The duration of the closed stochastic epidemic. *Biometrika*, 62(2): 477–482, 1975.

T. Bedford, S. Riley, I.G. Barr, S. Broor, M. Chadha, N.J. Cox, R.S. Daniels, C.P. Gunasekaran, A.C. Hurt, A. Kelso, A. Klimov, N.S. Lewis, X. Li, J.W.

McCauley, T. Odagiri, V. Potdar, A. Rambaut, Y. Shu, E. Skepner, D.J. Smith, M.A. Suchard, M. Tashiro, D. Wang, X. Xu, Lemey P., and C.A. Russell. Global circulation patterns of seasonal influenza viruses vary with antigenic drift. *Nature*, 523(7559):217–220, 2015.

J. Blanchet, P. Glynn, and S. Zheng. Empirical analysis of a stochastic approximation approach for computing quasi-stationary distributions. In *EVOLVE-A Bridge between Probability, Set Oriented Numerics, and Evolutionary Computation II*, pages 19–37. Springer, 2013.

P. Cattiaux, P. Collet, A. Lambert, S. Martínez, S. Méléard, and J. San Martín. Quasi-stationary distributions and diffusion models in population dynamics. *The Annals of Probability*, 37(5):1926–1969, 2009.

J.A. Cavender. Quasi-stationary distributions of birth-and-death processes. *Advances in Applied Probability*, 10(3):570–586, 1978.

Nicolas Champagnat, Amaury Lambert, and Mathieu Richard. Birth and death processes with neutral mutations. *International Journal of Stochastic Analysis*, 2012, 2012.

Y. Chen, J. Xie, and Jun S. Liu. Stopping-time resampling for sequential Monte Carlo methods. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(2):199–217, 2005.

N. Chopin. A sequential particle filter method for static models. *Biometrika*, 89(3): 539–552, 2002.

A. Cintrón-Arias, C. Castillo-Chávez, L.M. Bettencourt, A.L. Lloyd, and H.T. Banks. The estimation of the effective reproductive number from disease outbreak data. *Mathematical Biosciences and Engineering*, 6(2):261–282, 2009.

D. Clancy and S.T. Mendy. The effect of waning immunity on long-term behaviour of stochastic models for the spread of infection. *Journal of Mathematical Biology*, 61(4):527–544, 2010.

D. Clancy and P.K. Pollett. A note on quasi-stationary distributions of birth-death processes and the SIS logistic epidemic. *Journal of Applied Probability*, 40:821–825, 2003.

C.W. Clark, F.W.J. Olver, R.F. Boisvert, and D.W. Lozier. *NIST Handbook of Mathematical Functions*. Cambridge University Press, 2010.

P. Collet, S. Martinez, and J. San Martin. *Quasi-Stationary Distributions: Markov Chains, Diffusions and Dynamical Systems*. Springer, 2001.

B.J. Cowling, M.S.Y. Lau, L. Ho, S. Chuang, T. Tsang, S. Liu, P. Leung, S. Lo, and E.H.Y. Lau. The effective reproduction number of pandemic influenza: prospective estimation. *Epidemiology (Cambridge, Mass.)*, 21(6):842–846, 2010.

D. Crisan and A. Doucet. Convergence of Sequential Monte Carlo methods. Technical Report 381, Department of Engineering, University of Cambridge, 2000.

Gaofeng Da, Maochao Xu, and Shouhuai Xu. On the quasi-stationary distribution of SIS processes. *Probability in the Engineering and Informational Sciences*, 30 (4):622–639, 2016.

J.N. Darroch and E. Seneta. On quasi-stationary distributions in absorbing continuous-time finite Markov chains. *Journal of Applied Probability*, 4(1):192–196, 1967.

P. Del Moral. *Feynman-Kac Formulae: Genealogical and Interacting Particle Systems with Applications*. Springer-Verlag, New York, 2004.

P. Del Moral, A. Doucet, and A. Jasra. Sequential Monte Carlo samplers. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(3):411–436, 2006.

R. Douc and O. Cappé. Comparison of resampling schemes for particle filtering. In *Proceedings of 4th International Symposium on Image and Signal Processing and Analysis (ISPA)*, pages 64–69. IEEE, 2005.

A. Doucet and A.M. Johansen. A tutorial on particle filtering and smoothing: Fifteen years later. In *Handbook of Nonlinear Filtering*, pages 656–704. Oxford University Press, 2009.

A. Doucet, N. De Freitas, and N. Gordon. *Sequential Monte Carlo methods in practice*. Springer, 2001.

R. M. Dudley. Distances of probability measures and random variables. *Ann. Math. Statist.*, 39(5):1563–1572, 10 1968.

R. Durrett. *Probability: Theory and Examples*. Cambridge University Press, 2010.

P.A. Ferrari, H. Kesten, S. Martinez, and P. Picco. Existence of quasi-stationary distributions. a renewal dynamical approach. *The Annals of Probability*, 23(2): 501–521, 1995.

W.M. Fitch, R.M. Bush, C.A. Bender, and N.J. Cox. Long term trends in the evolution of H(3) HA1 human Influenza type A. *Proceedings of the National Academy of Sciences*, 94(15):7712–7718, 1997.

J. Geweke. Bayesian inference in econometric models using monte carlo integration. *Econometrica*, 57(6):1317–1339, 1989.

M. Girvan, D.S. Callaway, M.E.J. Newman, and S.H. Strogatz. Simple model of epidemics with pathogen mutation. *Physical Review E*, 65(3):031915, 2002.

J.R. Gog and B.T. Grenfell. Dynamics and selection of many-strain pathogens. *Proceedings of the National Academy of Sciences*, 99(26):17209–17214, 2002.

N.J. Gordon, D.J. Salmond, and A.F.M. Smith. Novel approach to nonlinear/non-Gaussian Bayesian state estimation. In *IEE Proceedings F (Radar and Signal Processing)*, volume 140 (2), pages 107–113. IET, 1993.

P. Groisman and M. Jonckheere. Simulation of quasi-stationary distributions on countable spaces. *Markov Processes and Related Fields*, 19(3):521–542, 2013.

M. Hairer, A.M. Stuart, and S.J. Vollmer. Spectral gaps for a Metropolis-Hastings algorithm in infinite dimensions. *Annals of Applied Probability*, 24(6):2455–2490, 2014.

S.D. Jacka and G.O. Roberts. Weak convergence of conditioned processes on a countable state space. *Journal of Applied Probability*, 32(4):902–916, 1995.

P.A. Jenkins. Stopping-time resampling and population genetic inference under coalescent models. *Statistical Applications in Genetics and Molecular Biology*, 11 (1):Article 9, 2012.

M. Kamo and A. Sasaki. The effect of cross-immunity and seasonal forcing in a multi-strain epidemic model. *Physica D: Nonlinear Phenomena*, 165(3):228–241, 2002.

W.O. Kermack and A.G. McKendrick. A contribution to the mathematical theory of epidemics. In *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering sciences*, volume 115, pages 700–721. The Royal Society, 1927.

J. F. C. Kingman. The exponential decay of Markov transition probabilities. *Proceedings of the London Mathematical Society*, s3-13(1):337–358, 1963.

R.J Kryscio and C. Lefèvre. On the extinction of the SIS stochastic logistic epidemic. *Journal of Applied Probability*, pages 685–694, 1989.

A. Lambert. Population dynamics and random genealogies. *Stochastic Models*, 24: 45–163, 2008.

J.S. Liu and R. Chen. Blind deconvolution via sequential imputations. *Journal of the American Statistical Association*, 90(430):567–576, 1995.

J.S. Liu and R. Chen. Sequential Monte Carlo methods for dynamic systems. *Journal of the American Statistical Association*, 93(443):1032–1044, 1998.

P. Mandl. On the asymptotic behaviour of probabilities within groups of states of a homogeneous Markov chain. *Časopis Pro Pěstování Matematiky*, (84):140–149, 1960.

P. Mandl. Spectral theory of semi-groups connected with diffusion processes and its application. *Czechoslovak Mathematical Journal*, 11(4):558–569, 1961.

J. Martins, A. Pinto, and N. Stollenwerk. Stationarity in moment closure and quasi-stationarity of the SIS model. *Mathematical Biosciences*, 236(2):126–131, 2012.

S. Méléard and D. Villemonais. Quasi-stationary distributions and population processes. *Probability Surveys*, 9:340–410, 2012.

S.C. Moy. Extensions of a limit theorem of Everett, Ulam and Harris on multitype branching processes to a branching process with countably many types. *The Annals of Mathematical Statistics*, 38(4):992–999, 1967.

I. Nåsell. On the quasi-stationary distribution of the stochastic logistic epidemic. *Mathematical Biosciences*, 156:21–40, 1999.

I. Nåsell. Stochastic models of some endemic infections. *Mathematical Biosciences*, 179(1):1–19, 2002.

P. Neal. Multitype randomized Reed-Frost epidemics and epidemics upon random graphs. *The Annals of Applied Probability*, 16(3):1166–1189, 2006.

H. Nishiura and G. Chowell. The effective reproduction number as a prelude to statistical estimation of time-dependent epidemic trends. In *Mathematical and Statistical Estimation Approaches in Epidemiology*, pages 103–121. Springer, 2009.

J.R. Norris. *Markov Chains*. Cambridge University Press, 1997.

P.D. O'Neill. Constructing population processes with specified quasi-stationary distributions. *Stochastic Models*, 23(3):439–449, 2007.

A. Parisi, J.S. Lopes, An. Nunes, and M.G.M. Gomes. Heterogeneity in antibody range and the antigenic drift of Influenza A viruses. *Ecological Complexity*, 14: 157–165, 2013.

L. Pellis, F. Ball, and P. Trapman. Reproduction numbers for epidemic models with households and other social structures. I: Definition and calculation of R_0. *Mathematical Biosciences*, 235(1):85–97, 2012.

P. K. Pollett. Reversibility, invariance and μ-invariance. *Advances in Applied Probability*, 20(3):600–621, 1988.

P.K. Pollett. Quasi-stationary distributions for continuous time Markov chains when absorption is not certain. *Journal of Applied Probability*, 36:268–272, 1999.

P.S. Puri. Interconnected birth and death processes. *Journal of Applied Probability*, 5(2):334–349, 1968.

I.W. Saunders. Epidemics in competition. *Journal of Mathematical Biology*, 11(3): 311–318, 1981.

M. Schotsaert and A. García-Sastre. A high-resolution look at Influenza virus antigenic drift. *Journal of Infectious Diseases*, page jiw183, 2016.

E. Seneta and D. Vere-Jones. On quasi-stationary distributions Markov chains with a denumerable infinity of states. *Journal of Applied Probability*, 3(2):403–434, 1966.

D. Sirl, H. Zhang, and P. Pollett. Computable bounds for the decay parameter of a birth-death process. *Journal of Applied Probability*, 44(2):476–491, 2007.

S.E.F. Spencer. *Stochastic Epidemic Models of Emerging Diseases*. PhD thesis, University of Nottingham, 2007.

S. Tavaré. Line-of-descent and genealogical processes, and their applications in population genetics models. *Theoretical Population Biology*, 26(2):119–164, 1984.

E.A. van Doorn. Quasi-stationary distributions and convergence to quasi-stationarity of birth-death processes. *Advances in Applied Probability*, 23(4):683–700, 1991.

E.A. van Doorn. Representations for the decay parameter of a birth-death process based on the Courant-Fischer theorem. *Journal of Applied Probability*, 52:278–289, 2015.

E.A. van Doorn and P.K. Pollett. Quasi-stationary distributions for reducible absorbing Markov chains in discrete time. *Markov Processes and Related Fields*, 15: 191–204, 1999.

E.A. van Doorn and P.K. Pollett. Survival in a quasi-death process. *Linear Algebra and its Applications*, 429:776–791, 2008.

E.A. van Doorn and P.K. Pollett. Quasi-stationary distributions for discrete-state models. *European Journal of Operational Research*, 230:1–14, 2013.

D. Vere-Jones. Some limit theorems for evanescent processes. *Australian Journal of Statistics*, 11(2):67–78, 1969.

World Health Organisation. Recommended composition of influenza virus vaccines for use in the 2015-2016 northern hemisphere Influenza season, 2015.

S. Wright. Evolution in Mendelian populations. *Genetics*, 16(2):97–159, 1931.

A.M. Yaglom. Certain limit theorems in the theory of branching processes. *Doklady Akademii Nauk SSR*, (56):795–798, 1947.

H. Zhang and W. Liu. Domain of attraction of the quasi-stationary distribution for the linear birth and death process. *Journal of Mathematical Analysis and Applications*, 385(2):677–682, 2012.